1973

Robustness of Homogeneity Tests in Parallel-Piped Contingency Tables

Raymond Joseph McNamee
Loyola University Chicago

Robustness of Homogeneity Tests In

Parallel~~e~~piped Contingency Tables

A Dissertation

Submitted to the Faculty of the Graduate School

of Loyola University

By

Raymond Joseph McNamee

In Partial Fulfillment of the Requirements

for the Degree

of

Doctor of Philosophy

September, 1973

ACKNOWLEDGEMENTS

The author is indebted to many friends and relatives for their assistance in reducing the data in this study; more than can be acknowledged by name. He is especially indebted to Professor Samuel T. Mayo for the basic ideas, his suggestions and encouragement. Without my son, Dennis, some of the tedious work would have taken six months longer to accomplish. Mike's humor injected into some trying situations was a form of encouragement. The music of Dick's bagpipes were a source of inspiration. Further aid was received from Mary Pat by her amusement of Sheila for many hours while she was being deprived of her rightful playtime with me, and from Sheila, her constant sunny smile.

To my wife, Irene, the author expresses gratitude for her constant encouragement, forbearance and whose typing made this dissertation appear better than it actually is.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The behavioral sciences are in the relatively early stages of development when compared to such sciences as physics and chemistry. Many studies in these evolving sciences are characterized by a search for fundamental variables.[1,2,3,4] Such a search is often one for general relationships and associations. One method for determining whether an association between such variables exists is to construct a multivariate frequency table, with a sample of N independent observations in which the categories of each variable are mutually exclusive and exhaustive. Tables of this kind are called contingency tables, and with such tables the chi-square test provides an appropriate test of independence. In applying chi-square to test for independence in an contingency table, the expected cell frequencies are derived from the data. These expected cell frequencies are those which would exist if the variables were

---

[1]J. M. Atthowe, "Behavior Innovation and Persistance", American Psychologist, 1973, 28, pp. 34-42.

[2]A. S. Cohan, "Career Patterns in the Irish Political Elite", British Journal of Political Science, 1973, 3, pp. 213-228.

[3]S. J. Morse, "Help, Likeability, and Social Influence", Journal of Applied Social Psychology, 1972, 2, pp. 24-34.

[4]A. M. Greeley, "The Religious Factor and Academic Careers", American Journal of Sociology, 1973, 78, pp. 1247-1256.

independent of each other and are different for first order and second order interaction. If the value of chi-square is considered significant at some accepted level, the null hypothesis that no difference exists between the observed and expected values is rejected. The alternative hypothesis that the two variables are associated may be accepted.

In a three dimensional contingency table the data are comprised of observations on three variables simultaneously. If the variables are dependent there may exist associations or interactions between these variables in pairs in which case first order interaction is said to exist. When interaction exists among the three variables, second order interaction is said to exist. Whether one or both types of interaction are of interest depends upon the relationships that are being investigated.

Technical literature related to the chi-square study of relationships among variables, describes those significant relationships as interactions and as associations. Darroch[5] states, "That interaction in contingency tables enjoys only a few of the fortuitously simple properties of interaction in

---

[5] J. M. Darroch, "Interactions in Multi-Factor Contingency Tables", Journal of Royal Statistical Society, Series B, 1962, 24, p. 254.

the analysis of variance" and Mayo[6] states "The word association is more often used for categorical variables." Thus it would have been better if authors had left interaction as a term specifically related to analysis of variance in order to avoid confusion. Most authors including Darroch use these terms interchangeable. Consequently in this paper the words interaction and association will be used synonymously.

The chi-square test on a contingency table knows no limitation as to the number of variables that can be studied simultaneously for association except for the practical limits of sample size. Although each variable adds a new dimension to the contingency table, the resulting patterns of interaction or association become more complicated while the amount of literature related to such contingency tables becomes increasingly sparse.

Another complexity associated with a chi-square test for interaction lies in the selection of small samples which directly relates to the sample size of the cells in the contingency table. The concern for small samples is due to the fact that in the derivation of the formula for the chi-square distribution, an

---

[6] S. T. Mayo, "Interactions Among Categorical Variables", Educational & Psychological Measurement, 21, No. 4, Winter, 1961, p. 839.

integral is substituted for a summation of discrete quantities. This approximation introduces an error that is of consequence when the sample sizes are small. The methodological question then becomes: how robust is the chi-square distribution for small samples?

Robustness plays an important role in statistical inference and methodology. Broadly speaking, a statistical procedure is said to be robust if it is insensitive to slight departures from postulated assumptions. These usually pertain to the parametric form of the underlying distribution and independence of the random variables. As the sample size increases, owing to the operation of the central limit theorem, many of the statistical tests become robust due to the operation of the central limit theorem. For example, the chi-square test in contingency tables becomes more robust as the total sample size increases. Thus it can be inferred that large sample situations do not normally pose a problem regarding robustness. However, an interesting methodological question often arises in this connection; namely, how large should the sample size be in order for the procedure to be "sufficiently" robust? Or, conversely, how small may a theoretical frequency be before it is considered too small?

Unfortunately, there are no simple answers to these problems.[7] When an exact answer is not readily available from an analytic study, then sometimes an empirical study will yield the answer with the desired accuracy which is a function of the number of computations made. A Monte Carlo study is such an empirical study where a random sampling is made of a simulated population on a computer.

Small cell sample sizes could still result in a large total sample size as in the case of a 2 x 2 x 12 contingency table. If a study necessitated the use of such a contingency table with a minimum cell sample of ten[8], a total sample size well in excess of 600 would be required. Whereas if the robustness of the chi-square test is such that the cell samples might be an average of four, then the total sample needed for such a study would be in the vicinity of 250.

---

[7]G. U. Yule, and M. G. Kendall, An Introduction to the Theory of Statistics, (London: Griffin, 1947), p. 422.

[8]D. Lewis and C. J. Burke, "The Use and Misuse of the Chi-Square Test", Psychological Bulletin, 1949, 46, p. 456.

The study will attempt to focus on three questions associated with the chi-square tests for first and second order interaction in 2 x 2 x d contingency tables.

(1) Does the magnitude and direction of the calculated chi-square for first and second order interaction vary with table sample size and dimension of the contingency table?

(2) Is the average table sample size or the 2 x 2 subtable sample size the determining factor of the robustness of the chi-square tests for first and second order interaction?

(3) How robust is Norton's iteration technique for determining the chi-square for second order interaction?

STATEMENT OF THE PROBLEM

Lewontin and Felsenstein[9] have shown that the chi-square test for homogeneity is quite robust for small samples in two dimensional contingency tables. In order to enable research personnel to effectively utilize three dimensional contingency tables with small sample cells, a determination of the robustness of homogeneity tests in these tables is necessary. The data in three dimensional contingency tables cannot be analyzed by pooling the data and forming a series of two dimensional tables has been shown by Simpson,[10] Fricke,[11] and Kullback.[12] This study will determine the robustness of homogeneity tests on contingency tables of dimensions 2 x 2 x d where d will take the values between 2 and 12. The tests under consideration are the chi-square tests for first and second order interaction. Since only the upper tail of the chi-square

---

[9] R. C. Lewontin and J. Felsenstein, "The Robustness of Homogeneity Tests in 2 x n Tables", Biometrics, 1968, 24, pp. 19-33.

[10] E. H. Simpson, "The Interpretation of Interaction in Contingency Tables", Journal of Royal Statistical Society, Series B, 1951, 13, pp. 238-241.

[11] B. G. Fricke, "A Configural-Content-Intensity Item For Personality Measurement", Educational and Psychological Measurement, 1956, 16, pp. 54-62.

[12] S. Kullback, Information Theory and Statistics, New York, Dover, 1968, p. 343.

distribution where the cumulative distribution exceeds 0.90 is important for testing the null hypothesis, agreement in this region of the distribution would have the effect of making the chi-square tests robust, irrespective of the correspondence between these distributions for smaller values of chi-square.

The variables that will be manipulated in this study are as follows:

(1) the sample size of each 2 x 2 sub-contingency table

(2) the number of categories of one variable (d) in the three dimensional tables.

This chapter has attempted to briefly outline the problem. The following chapter contains a review of the literature concerning contingency tables. Chapter Three contains a description of the design and methods used in the investigation. The results of the investigation and their interpretation are described in detail in Chapter Four. The fifth and last chapter contains the summary, conclusions and suggestions for further research. The computer routines that were used in the study are contained in the Appendix.

CHAPTER II

Review of Related Literature

Introduction

The opening sentence of the initial work in the area of analysis of higher order interaction in contingency tables should be noted here. Its importance is accentuated since not only is Bartlett the first to have made the observation; but the statement is still true today, nearly forty years later:

> Although the structure and analysis of ordinary contingency tables with two-way classifications have been subjected to much critical examination, the complex table involving more than two ways of classifying the data seems, perhaps because it is rather unusual, to have been comparatively ignored.[1]

Equally as significant are the statements of Mayo and of Lewis some twenty-five years later in relation to the same issue. Mayo states for example:

> When it comes to the case of contingency tables with three or more attributes, the textbooks are even barer. There is some periodical literature on higher order interactions, but one must look through many widespread sources to get a meaningful picture.[2]

---

[1]M. S. Bartlett, "Contingency Table Interactions", Journal of Royal Statistical Society Supplement, 1935, 2, p. 248.

[2]S. T. Mayo, "Interactions Among Categorical Variables", Educational & Psychological Measurement, Vol. 21, No. 4, Winter 1961, p. 840.

Lewis, commenting on the same issue, states:

> It is now over twenty-five years since Bartlett first commented on the lack of interest shown in contingency tables of three or more dimensions. Since then several important papers have appeared on the subject, but there is still no coordinated information available, and the treatment of these tables is still widely neglected in standard textbooks.[3]

One may assume that the reason the analysis of complex contingency tables is so rare is the mathematical difficulty in analyzing them, and perhaps the fact that the different techniques of testing higher order interactions do not always give comparable results.[4]

In this chapter an examination of the literature concerning the definition of higher-order interaction, degrees of freedom for the chi-square tests and the effects of sample size will be made.

The attempts to arrive at a logical, consistent, and intuitively acceptable definition of higher order interaction

---

[3]B. N. Lewis, "On the Analysis of Interaction in Multidimensional Contingency Tables", Journal of Royal Statistical Society, Volume 125, 1962, p. 88.

[4]Mayo, loc. cit.

as well as the resulting techniques of analyzing complex contingency tables can be divided into the following four classifications:

(1) Bartlett's original definition and its extensions by Norton, Roy and Kastenbaum, and later by Darroch.

(2) Formulations based on the symmetrical functions of the cell probabilities by Simpson and Plackett.

(3) Definition of Interaction based on maximum entropy.

(4) Definition of Interaction based on information theory.

An examination of these four attempts to explain higher order interaction follows:

1. <u>Bartlett's Definition and its Extensions</u>

Bartlett[5] was the first to formulate a definition for second order interaction. He formulated his hypothesis, proposed the statistic and suggested a method for the solution in less than twenty-five lines. His formulation for second order interaction in a 2 x 2 x 2 table is an extension of the formulation of first order interaction in

---

[5]Bartlett, op. cit., pp. 248-252.

a 2 x 2 table. It depends on the assumption that the cross product ratio hypothesis could be extended to define second order interaction for 2 x 2 x 2 tables. Without stating what he considered to be the dividing line between small samples and large samples he developed both a small sample theory and a large sample theory. Finally, he showed how his definition could be extended to a 2 x 2 x 3 table and the resulting pair of simultaneous cubic equations.

From 1935 to 1945 the problem of second order interaction remained in a kind of an intellectual limbo. The next step was taken on this side of the Atlantic by H. W. Norton,[6] who developed an iterative technique for solving the simultaneous cubic equations resulting from Bartlett's definition of second order interaction in 2 x 2 x n contingency tables.

The extension of Norton's method to three dimensional r x s x t tables was derived theoretically in a paper by S. N. Roy and Marvin A. Kastenbaum.[7] The theoretical basis

---

[6]H. W. Norton, "Calculation of Chi-Square for Complex Contingency Tables", Journal of American Statistical Association, 1945, 40, pp. 251-258.

[7]S. N. Roy and M. A. Kastenbaum, "On the Hypothesis of No Interaction in a Multiway Contingency Table", Annals of Mathematical Statistics, 1956, 27, pp. 749-757.

for their paper was a series of mimeographs from the Institute of Statistics at the University of North Carolina which were written by S. K. Mitra,[8] S. N. Roy and M. Kastenbaum.[9, 10] (Copies of these mimeographs are at the Mathematical Library of the University of Chicago.) A paper dealing with the application of the Roy and Kastenbaum formulations to the calculation of chi-square to test no three-factor interaction in multicategory three dimensional contingency tables was published in 1959 by Kastenbaum and Lamphiear.[11] This paper also notes the coming of age of the digital computer.

---

[8] S. K. Mitra, "Contributions to the Statistical Analysis of Categorical Data", (Institute of Statistics, University of North Carolina, Mimeograph Series No. 142, 1955.)

[9] S. N. Roy and M. Kastenbaum, "A generalization of analysis of variance and multivariate analysis to data based on frequencies in qualitative categories or class intervals." (Institute of Statistics, University of North Carolina, Mimeograph Series No. 131, 1955.)

[10] S. N. Roy and S. K. Mitra, "An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis." (Institute of Statistics, University of North Carolina, Mimeograph Series No. 139, 1955.)

[11] M. A. Kastenbaum and D. E. Lamphiear, "Calculation of Chi-Square to Test the No Three Factor Interaction Hypothesis", Biometrics, 1959, 15, pp. 107-115.

It is the purpose of this paper to demonstrate a technique which, while practical with an ordinary desk calculator, is particularly well suited for modern high-speed computers.[12]

Darroch[13] considered his formulation of a hypothesis of no second order interaction in a three way contingency table as a direct continuation of the work of Kastenbaum and Lamphiear. He made an explicit comparison of the definitions of interaction in multiway contingency tables and in the analysis of variance. His likelihood ratio test was found to be asymptotically distributed as chi-square with (r-1) x (c-1) x (d-1) degrees of freedom. Darroch's paper is an intellectual bridge from papers in this classification to papers utilizing symmetrical functions of the cell probabilities. He defines a perfect three-way table to which a symmetrical definition of no second order interaction

---

[12]Ibid, p. 107.

[13]J. N. Darroch, "Interaction in Multi Factor Contingency Tables", Journal of Royal Statistical Society, Series B, 1962, 24, p. 252.

would apply. A symmetrical definition of no second order interaction with respect to the three classifications of a 2 x 2 x 2 table would be as follows. If some function $F(p_{111}, p_{121}, p_{211}, p_{221})$ is chosen as a measure of the association of classifications R and C in D, then the function must be such that the equation

$$F(p_{111}, p_{121}, p_{211}, p_{221}) = F(p_{112}, p_{122}, p_{212}, p_{222})$$

which implies and is implied by the equations

$$F(p_{111}, p_{211}, p_{112}, p_{212}) = F(p_{121}, p_{221}, p_{122}, p_{222})$$

and

$$F(p_{111}, p_{121}, p_{112}, p_{122}) = F(p_{211}, p_{221}, p_{212}, p_{222}).$$

2. Formulations Based on the Symmetrical Functions of the Cell Probabilities

In 1951 Simpson[14] defined no second order interaction to be symmetrical with respect to the three attributes of a three dimensional contingency table. He showed that the cross product ratio used by Bartlett satisfies this requirement and thus reached Bartlett's conslusion by an alternate approach. The editor in a footnote to Simpson's paper suggested that this paper should be read in conjunction with the following paper by H. O. Lancaster. Lancaster[15] defined second order interaction as the difference between the total chi-square statistic for testing complete independence of the three classifications, and the sum of the three components corresponding to tests for independence in each of the three marginal tables. It was not until 1962

---

[14] E. H. Simpson, "The Interpretation of Interaction in Contingency Tables", Journal of Royal Statistical Society, Series B, 1951, 13, p. 239.

[15] H. O. Lancaster, "Complex Contingency Tables Treated by the Partition of Chi-Square", Journal of Royal Statistical Society, Series B, 1951, 13, pp. 242-249.

that Plackett[16] compared these two definitions, showing that the latter does not always satisfy the condition of symmetry. Plackett accepted Roy's and Kastenbaum's definition for a three dimensional table and extended the analysis of log-frequencies to such tables as an alternative method of analysis which is computationally easier than the solution of (r-1) x (c-1) x (d-1) simultaneous equations of the third degree. However, he noted that

> "The (log-frequency) test described above is, of course, valid only when n approaches infinity."[17]

---

[16] R. L. Plackett, "A Note on Interactions in Contingency Tables", Journal of Royal Statistical Society, Series B, 1962, 24, pp. 162-166.

[17] Ibid., p. 166.

3. Definition of Interaction Based on Maximum Entropy

The formulations of the hypothesis of no second order interaction summarized up to now are basically extensions of Bartlett's work. Good[18] proposed to use the principle of maximum entropy as a heuristic principle for the generation of interaction to that of no rth-order and all higher order interactions in an m-dimensional contingency table with a complete set of rth-order restraints by means of discrete Fourier transforms of the logarithms of probabilities. However, the interactions so defined are usually complex valued unless the categories within each classification are equal to two.

Goodman[19] followed the definition by Good but proposed a test that yields real valued interactions. His proposed test was based on Wald's criterion and unrestricted maximum likelihood estimates which was essentially an extension of the tests proposed by Plackett.

---

[18] I. J. Good, "Maximum Entropy for Hypothesis Formulation Especially for Multidimensional Contingency Tables", Annals of Mathematical Statistics, 1963, 34, p. 922.

[19] E. A. Goodman, "On Plackett's Test for Contingency Table Interactions", Journal of Royal Statistical Society, Series B, 1963, 25, p. 187.

While Good's and Goodman's formulations and tests of no interaction hypotheses are entirely general, physical interpretations of their meanings became extremely difficult for interactions higher than the second order. In the case of second order interactions, their formulations reduce to the ones of Bartlett, Norton, Roy and Kastenbaum, and Darroch.[20]

These papers by Good and Goodman form a bridge of ideas between the previous formulations of no interaction and the information theoretic approach by introducing the idea of entropy which, in this context, is related to "missing information".[21] This is true since entropy is a function of alternatives which remain possible to a system after all the macroscopically observable information concerning it has been recorded.

---

[20]Ibid.

[21]Claude E. Shannon and Warren Weaver, The Mathematical Theory of Communication (Urbana: University of Illinois Press, 1963), p. 95.

4. Definition of Interaction based on Information Theory

The information theory school of thought, on the determination of interaction in multidimensional contingency tables, is represented by H. H. Ku, S. Kullback, M. Kupperman and C. J. Ireland. In general, information measures are applicable whenever data can be arranged in the form of a contingency table because if two or more of the classifications interact, each can be thought of as providing information reducing the uncertainty about the other classifications. In the case of a three dimensional contingency table, both the row and column classifications can contribute information (jointly or independently) about the layer classification. By procedures similar to analysis of variance the total amount of information can be partitioned into additive components associated with individual classifications and their interactions. In his exposition of the above, Kullback[22] shows that many of his information measures are almost numerically identical to the chi-square measures of interaction.

---

[22]Solomon Kullback, Information Theory and Statistics, New York, Dover, 1968, 8, pp. 155-158.

In 1968 Ku and Kullback[23] published a summary type paper in which they used an information theory approach to compute the minimum discrimination information statistic (m.d.i.s.). They point out that it is equivalent to minus twice the natural logarithm of the likelihood ratio. Their definition of no second order interaction is shown to be equivalent definitions of no second order interaction in a 2 x 2 x 2 table given by Bartlett, no interaction in an r x s x t table by Roy and Kastenbaum; it is also related to that definition given by Good.

All the authors of the last three approaches to a higher order interaction in a contingency table justify their algorithm by showing agreement with the definition of Bartlett and its extension by Roy and Kastenbaum. Norton's paper is basically an exposition of an algorithm for solving the cubic equations resulting from Bartlett's definition. The reason the various authors have given for formulating their respective algorithm from alternate definitions was to reduce the amount of labor involved in

---

[23]H. H. Ku and S. Kullback, "Interaction in Multidimensional Contingency Tables: An Information Theoretic Approach", <u>Journal of Research of the National Bureau of Standards, Mathematical Sciences</u>, 1968, 72B, p. 166.

calculating the probability of second order interaction by Bartlett's definition. However, in 1959, Mayo made a prophetic statement:

> With the advent of electronic computers, one should find that older, formerly prohibitive techniques will become feasible and probably newer computing programs will become available.[24]

With the IBM 370 and Fortran IV, Norton's iteration technique is not only feasible but quite straightforward. Therefore, since none of these authors have found anything theoretically wrong with Bartlett's definition and Norton's algorithm and, as a matter of fact, have used them as standards in determining the accuracy of their algorithms, Norton's algorithm will be used in this dissertation.

---

[24] S. T. Mayo, "Towards Strengthening the Contingency Table as a Statistical Method", Psychological Bulletin, 1959, 56, p. 470.

Degrees of Freedom

> Not only do most texts omit mention of the concept but many actually give incorrect formulas and procedures because of ignoring it.[25]

There is an apparent disagreement concerning the number of degrees of freedom for the first order interaction chi-square in three dimensional contingency tables. For example, Conover[26] and Pierce[27] in their respective books, claim the number of degrees of freedom should be $(r-1) \times (c-1) \times (d-1)$. Lewis[28] and Kullback[29] state that the number of degrees of freedom should be $rcd-r-c-d+2$. Obviously, both cannot be right. In Walker's[30] development of degrees of freedom in a two way contingency table, it was shown that in the two way table the number of degrees of freedom would be $rc-r-c+1$. Extending her reasoning to a

---

[25] H. M. Walker, "Degrees of Freedom", Journal of Educational Psychology, 1940, 31, p. 253.

[26] W. J. Conover, Practical Nonparametric Statistics, New York: Wiley, 1971, p. 165.

[27] A. Pierce, Fundamentals of Nonparametric Statistics, New York: McGraw-Hill, 1970, p. 214.

[28] B. N. Lewis, "On the Analysis of Interaction in Multidimensional Contingency Tables", Journal of Royal Statistical Society, Series 1962, 125, p. 90.

[29] Solomon Kullback, Information Theory and Statistics, New York: Dover, 1968, p. 165.

[30] H. M. Walker, loc. cit., p. 265.

three way contingency table the number of degrees of freedom would be rcd-r-c-d+2, which would reduce to rc-r-c+1 when d equals 1. The discrepancy between the number of degrees of freedom being (r-1) x (c-1) x (d-1) or rcd-r-c-d+2 for the three dimensional contingency table may have arisen from the fact that in the two way contingency table the number of degrees of freedom being equal to rc-r-c+1 can be written in the factored form (r-1) x (c-1) while rcd-r-d+2 does not factor into (r-1) x (c-1) x (d-1).

All authors who discuss second order interaction agree that the number of degrees of freedom for second order interaction in a three dimensional contingency table is (r-1) x (c-1) x (d-1). In Pierce's[31] book, there is a computer program in the BASIC language for computing the first order interaction chi-square statistics in a three dimensional table. It uses (r-1) x (c-1) x (d-1) for determining the number of degrees of freedom.

---

[31]A. Pierce, loc. cit., pp. 290-294.

SAMPLE SIZE

A study with a large number of well-controlled observations will give results in which more confidence can be placed than a study with a small number of equally well-controlled observations. However, sometimes a "trade-off" situation occurs where an increase in the number of observations either decreases the controls or appreciably increases the cost. The question then becomes - "How small may a frequency be before it is considered too small to yield valid conclusions?"

The problem of the distribution of the calculated statistic from the rectangular contingency table, and its relation to the theoretical chi-square distribution when the expectations are small has received considerable attention, and various investigations have been reviewed and summed up by Cochran.[32,33] In these articles, Cochran gives a number of rules of thumb for choosing the proper tests.

---

[32] W. G. Cochran, "The Chi-Square Test of Goodness of Fit", Annals of Mathematical Statistics, 1952, 23, pp. 315-345.

[33] W. G. Cochran, "Some Methods for Strengthening the Common Chi-Square Test", Biometrics, 1954, 10, pp. 417-451.

These rules are based on a number of special cases and his experience with a variety of models. How small does a sample have to be before it is considered small? According to Fisher[34] in his Statistical Methods, five is the number below which the expectation values should not fall if the chi-square test is to be used. Lewis and Burke[35] believe ten is the lower limit. Cochran[36] suggests that the chi-square test is a good approximation for a 2 x n contingency table when all expected values are at least two and when most cells have an expected value of five or more. In the present study a few cells may even have expected values of one without serious inaccuracy. With small samples one can apply Yates' correction[37] for continuity in contingency tables with one degree of freedom, but even here there is

---

[34] R. A. Fisher, Statistical Methods for Research Workers, (Edinburgh: Oliver and Boyd, 1936), p. 83.

[35] D. Lewis and C. J. Burke, "The Use and Misuse of the Chi-Square Test", Psychological Bulletin, 1949, 46, pp. 433-489.

[36] Cochran, loc. cit.

[37] F. Yates, "Contingency Tables Involving Small Numbers of the Chi-Square Test", Journal of Royal Statistical Society, Suppl. 1, 1934, pp. 217-235.

debate as to when this correction overcompensates and biases the test toward higher probabilities.

In practice, as Fisher[38] states with his usual common sense attitude that interest is not primarily in the exact value of the probability that the distribution of frequencies is due to a given hypothesis, but rather in whether or not the hypothesis is open to suspicion. Even though Lewis and Burke, in their article "The Use and Misuse of the Chi-Square Test", placed the most stringent requirements concerning sample size of all the authors reviewed, they said in another article:

> In many cases chi-square tests based on small theoretical frequencies are not far off, and there appears to be the possibility of proving that such tests will never be far off. Until such proof is forthcoming, however, the possibility must be accepted that combinations of numbers in contingency tables may exist for which the divergencies could not be neglected. In this connection, we will be quite happy to see final proof that we are wrong.[39]

---

[38]R. A. Fisher, Statistical Methods for Research Workers, (Edinburgh: Oliver and Boyd, 1936), p. 83.

[39]D. Lewis and C. J. Burke, "Further Discussion of the Use and Misuse of the Chi-Square Test", Psychological Bulletin, 1950, 47, p. 355.

There are two types of proof by induction. One is called Mathematical Induction and the other is Scientific Induction. Mathematical induction is a very limited type proof in that it works only where items may be placed into a one-to-one correspondence with the set of natural numbers. However, mathematical induction is never disproven at some later date. Many things proven by scientific induction on the other hand are later proven to be wrong when more facts concerning the situation are discovered. However, proofs by scientific induction have aided sciences like Chemistry and Physics in the advances they have made in the last few centuries. If Lewis and Burke[40] look upon Statistics as a branch of Mathematics, then the proof they are looking for in the last quote may still remain a long way off. On the other hand, if they look upon Statistics as a science that is aided by mathematics, then a proof concerning small cell sample sizes in a limited number of different types of rectangular contingency tables may be found in the articles by Lewontin and Felsenstein, and Craddock and Flood.

---

[40]Lewis and Burke, ibid.

Lewontin and Felsenstein summarize their results as follows:

> A Monte Carlo investigation of 2 x n tables with fixed marginals has been performed. The results of the Monte Carlo distribution show that the probability of Type I error given by the conventional chi-square test is in general conservative for five or more degrees of freedom even when expectations of successes are very small in each cell. A very conservative rule of operation would be that if expectations are one or greater the test is certainly conservative at the 5%, 2% and 1% levels of significance and that for most cases even fractional expectations do not affect the test.[41]

Craddock and Flood state:

> Our field of investigation differs from that of Cochran (1954) since he considered cases in which the expectancy in one cell of the contingency table fell below five whereas in our experiments all cell expectancies do so[42]...The results of this work have already proved useful in meteorological research and should be equally useful when similar categorized data arise elsewhere. It allows the chi-square statistic to be used under conditions in which it does not conform to the chi-square distribution.[43]

---

[41]R. C. Lewontin and J. Felsenstein, "The Robustness of Homogeneity Tests in 2 x n Tables", Biometrics, 1968, 24, p. 19.

[42]J. M. Craddock and C. R. Flood, "The Distribution of the Chi-Square Statistic in Small Contingency Tables", Journal of the Royal Statistical Society, Series C, 1970, 19, p. 173.

[43]Ibid, p. 181.

In both of these papers the experimental procedure that was used was a Monte Carlo technique. These authors have shown that the chi-square test is robust for rather small expected values. These two articles comprise the total of Monte Carlo investigation of small sample contingency tables. In a search of the literature no Monte Carlo investigations of higher order contingency tables were found.

CHAPTER 3

Design of the Study

Introduction

This chapter discusses the methodology, as well as the underlying theoretical basis, used to study the robustness of the chi-square test for first and second order interaction in 2 x 2 x n contingency tables. In this discussion the following ten points will be covered: 1) the rationale for using the Monte Carlo approach and necessary conditions, 2) the Monte Carlo procedure for determining robustness of the chi-square test, 3) the generation of the cell frequencies in the contingency tables to be examined, 4) an explanation of the notation used in the derivations of the first and second order expected values, 5) the rationale for the number of iterations, 6) how the 2 x 2 subtable sample sizes were determined, 7) how the chi-square statistic for first order interaction was determined, 8) how the chi-square statistic for second order interaction was determined for contingency tables with no zero valued cells, 9) the determination of the value of the chi-square statistic for second order interaction when the contingency table contains zero valued cells, and 10) a description of the real data to be examined as a demonstration of the usefulness of this study to the practitioner.

Rationale of the Study

In order to determine the robustness of the chi-square statistic in determining first and second order interaction in papallelepiped contingency tables, either analytic or Monte Carlo methods can be applied. Analytic methods would require a great deal of theoretic endeavor in the areas of multi-dimensional distributions, combinatorial analysis and the resulting multinomial coefficients. The chi-square distribution function is derived from the multinomial distribution function by the following approximations:[1]

1) replacing factorials by their Stirling approximation

2) replacing factors of the form $(1 + (X/n))^n$ with e^x when n is large

3) substituting a continuous integral for a summation of discrete quantities.

Therefore, to determine the robustness of the chi-square test by analytic methods would be to justify mathematically these approximations above and beyond the 90% level.

---

[1]D. Lewis and C. J. Burke, "The Use and Misuse of the Chi-Square Test", Psychological Bulletin, 1949, 46, pp. 433-489.

The results of such a study would possibly produce results beyond the bounds of this study, but these results would not be extremely useful to the educational practitioner. Conversely, Monte Carlo techniques are more easily implemented in this particular problem but will yield results that are not as general as analytic methods. However, important specific results concerning the robustness of the chi-square statistic for small sample contingency tables can certainly be discovered by this method.

Since Monte Carlo techniques are to be utilized for this study, it is necessary to establish some rationale by which these techniques will be employed. The selection of conditions and methodology will depend upon limitations of computing and operating time, and, in turn, on 1) the extent to which the situation might actually occur in research, and 2) particular combinations of factors procuding the most relevant and important information with respect to the problem viewed as a function of the research previously conducted in related areas. The points establish a basis that will be either directly or indirectly applied to provide rationale for the selection of procedures and conditions used in this study.

The Monte Carlo Procedure

Monte Carlo methods are distinguished by their experimental nature. Whether or not a particular experimental method is eligible to be called a Monte Carlo method may be largely a matter of personal taste. Monte Carlo methods are not known for all problems nor do specific problems necessarily admit a unique specific Monte Carlo approach. On the contrary, there may exist different Monte Carlo methods for a given problem, not obviously related one to another.

For definiteness, we shall class as a Monte Carlo method any procedure which involves the use of random numbers in obtaining samples to approximate the solution of a problem in the physical, social, and biological sciences.

This Monte Carlo study has two purposes. It seeks to determine if there is bias in the chi-square statistic for first and second order interaction when the cell sample sizes are between 3 and 12 and the stability of these statistics when the sample sizes of the 2 x 2 subtables are between 16 and 44. This information may then be used to determine correction factors for the chi-square distribution when the sample size is small.

Through the use of the computer (IBM 370) a 2 x 2 x d contingency table will be generated where "d" will vary between 2 and 12. The chi-square statistics for first and second order interaction will be calculated for each contingency table produced. This procedure will be repeated for t iterations for different values of cell expectation and contingency table length. Then the accumulated calculated values will be compared with the theoretical values for the 10%, 5%, 2%, and 1% levels of significance. The computer program was validated by hand calculation for five sample tables of different dimensions. The results of the computer program were in perfect agreement with Norton's results for the housefly problem.

<u>The Generation of Data</u>

A Monte Carlo program performed all computations utilizing the IBM 370 computer. This computer is the property of Triton College[2]. The program utilized for the study was written in Fortran IV and is given in the appendix.

A highly critical routine used in the program was the pseudo-random number generator RANDU. This subroutine generates a uniformly distributed pseudo-random number in the interval 0 to 1. The quality of the program has been closely examined by Pritsker and Kiviat[3], It is a subroutine of the Gasp II Simulation program developed at Arizona State University. These random numbers are then used to fill the cells of each 2 x 2 subtable, where each subtable has a fixed expectation value.

---

[2]Triton College, 2000 Fifth Avenue, River Grove, Illinois, 60171.

[3]A. Alan B. Pritsker and Philip J. Kiviat, <u>Simulation With Gasp II</u>, (Englewood Cliffs, New Jersey: Prentice Hall, 1969)

Each subtable is filled in the following manner. The random number generated by the subprogram is a number between 0 and 1. This number is multiplied by ten and the resulting first digit is retained. If this digit is greater than three, it is discarded and another random number is generated. When the resulting digit is three or less, one is added to the number and a frequency count is added to the cell labeled as follows:

0	1
2	3

This process is repeated until each subtable has acquired its designated amount. For example, consider the following random numbers and the resulting digits:

Random Number	x 10	Leading Digit	
0.013475	0.13475	0	
0.123578	1.23578	1	
0.357965	3.57965	3	
0.468024	4.68024	4	discard
0.140428	1.40428	1	
0.235680	2.35680	2	
0.003456	0.03456	0	
0.134567	1.34567	1	
0.897643	8.97643	8	discard
0.345678	3.45678	3	

The resulting subtable would have the following count in each cell:

2	3
1	2

Notation

Since it is common practice to use Roman letters to stand for sample statistics, C^2 will be used to designate the statistic associated with the chi-square parameter.

In the case of the three dimensional 2 x 2 x d contingency table with two rows, two columns, and d "layer" categories, let n_{ijk} denote the observed frequency in the cell of the ith row, jth column and kth layer. Similarly, let p_{ijk} denote the probability of an observation falling in the (ijk)th cell.

If the n_{ijk} frequencies are summed over both values of i, the result will clearly be the "marginal" total of the jth column in the kth layer. This marginal total is accordingly designated $n_{.jk}$, so that

$$n_{1jk} + n_{2jk} = n_{.jk} \text{ by definition.}$$

Similarly, $n_{i1k} + n_{i2k} = n_{i.k}$

and $$\sum_{k=1}^{k=d} n_{ijk} = n_{ij.}$$

The sum of n_{ijk} frequencies over all values of both i and j with k held constant yields the total of the kth layer. This total is designated $n_{..k}$ with $n_{i..}$ and $n_{ij.}$ defined in a similar manner.

Finally, if the n_{ijk} frequencies are summed over all values of i, j and k, the result will be the grand total of the contingency table. This is sometimes designated n... but is more conveniently written as N.

A similar notation is used for the probabilities (p's) except that the grand total probability (p...) is equal to one.

An example of the use of this notation with actual data[4] is as follows:

				C			
		1				2	
		B				B	
		1	2			1	2
A	1	112 (n_{111})	70 (n_{121})	A	1	144 (n_{112})	84 (n_{122})
	2	75 (n_{211})	110 (n_{221})		2	108 (n_{212})	133 (n_{222})
$n_{.jk}$		187 ($n_{.11}$)	180 ($n_{.21}$)			252 ($n_{.12}$)	217 ($n_{.22}$)
$n_{..k}$		367 ($n_{..1}$)				469 ($n_{..2}$)	
$n_{...}$				836			

[4]S. T. Mayo, "Interactions Among Categorical Variables", Educational & Psychological Measurement, 21, No. 4, 1961, p. 853.

The Number of Iterations

The function of the Monte Carlo routine as applied in the present study is to determine the proximity of the probability distribution, obtained from the calculated chi-square statistic of a random sample of contingency tables consisting of small samples, to the theoretical chi-square distribution above the ten percent level of significance. The number of iterations t will determine the precision of the estimates of the chi-square probability distribution.

In order to determine N, a procedure used by J. A. Kavanagh[5] will be utilized. Let C^2 be the statistic calculated from a contingency table by the following formula;

$$C^2 = \sum \frac{(\text{observed-expected})^2}{\text{expected}} ;$$

and $X^2_{.90}$ denote the 90th percentile of the theoretical chi-square distribution. Now consider the following:

$$y = \begin{cases} 1 \text{ for } \left\{ C^2 \mid C^2 \geq \chi^2_{.90} \right\} \\ 0 \text{ for } \left\{ C^2 \mid C^2 < \chi^2_{.90} \right\} \end{cases}$$

---

[5] J. A. Kavanagh, A Monte Carlo Study of the Polynomial Discriminant Method for Pattern Recognition, unpublished Ph.D Thesis, University of Minnesota, 1972, p. 26.

Consider t independent iterations or observations of y. If p is the probability that $C^2 \geq \chi^2_{.90}$,

then $\sum_{i=1}^{t} y_i$ is binomially distributed with parameter p

and $y = \sum_{i=1}^{t} y_i$ is an estimate of p.

By applying the central limit theorem to approximate the distribution of $\bar{y}$ as t gets larger, it is found that $\bar{y}$ is approximately normally distributed with a mean of p and variance of $\frac{p\,(1-p)}{t}$. From this the following probability statement can be made where Z_α is found in the standard normal tables.

$$P\left(\frac{\bar{y} - p}{\sqrt{\frac{p\,(1-p)}{t}}} \geq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

For $\alpha = .05$, we find $Z_{1-\frac{\alpha}{2}} = 1.96$.

The worst situation concerned with in this study is one in which p = .9 and hence the variance of y is equal to

$$\frac{(.9)\ (.1)}{t}\ .$$

If one is satisfied with an estimate of p that is within .03 of the actual value of p 95% of the time, then t is 385.

Also with t = 385, 95% of the time the precision of p = .95 is within .02 while the prècision of p = .99 is within .009.

If it were necessary to increase the precision of the estimate of p = .90 to .02 for 95% of the time the required t would be 3458. Since time and money are two important considerations in this study, t will be taken as 400.

Sub-Table Sample Sizes

The determining factor for the smallest sample for each 2 x 2 subtable was how often a zero would appear in a cell due to a random process.

In each 2 x 2 subtable there are four cells and the probability that a given cell will not receive a count for a given random number if 3/4 or 0.75. If the subtable sample size is 12 then the probability that a given cell will not receive a count for 12 random numbers is $(3/4)^{12}$. The probability that any one of the four cells of the subtable is zero-valued would be $4(3/4)^{12}$ or 0.1267 with four place accuracy. Hence the probability of obtaining a subtable without a zero would be 0.8733. Since there are 12 of these subtables in a 2 x 2 x 12 contingency table, the probability of obtaining a 2 x 2 x 12 contingency table with no zeros would be $(0.8733)^{12}$ or .1967. Therefore, fewer than one-fifth of the contingency tables would be used in predicting the value of the second order chi-square. If the subtable sample size is increased to 16 more than three-fifths of the 2 x 2 x 12 contingency tables would not contain a zero. However, if the sample size is decreased to 8 only two 2 x 2 x 12 tables in a thousand would contain all non-zero cells.

Therefore, the smallest sample for each 2 x 2 subtable that could be used to obtain measurable results would be 12 or an expected cell size of 3. This is below the minimum value of five discussed in Chapter Two. The upper level of the expected cell size was chosen as 12 which yields a subtable sample size of 48. Since ten was the greatest lower bound of sample sizes, discussed by the authorities quoted in Chapter Two for which the chi-square test was valid, a few values greater than ten were necessary in order to generate contingency tables whose average cell size was ten, in order to determine the stability of the statistics.

The stability of the chi-square statistic is determined by comparing the calculated values of chi-square when:

1) the subtable samples are allowed to vary between the smallest subtable sample size and the largest subtable sample size under consideration

2) the sum of each such combination of subtable samples is equal to N.

The following contingency tables will be analyzed. At the top of each list is the mean of the 2 x 2 subtable sample sizes listed horizontally. Vertically are all the combinations from the following sets of 2 x 2 subtable sample sizes (12, 16, 20, 24, 28, 32, 36, 40, 44, 48), that yield the listed mean. These subtables correspond to the following set of cell expected values (3, 4, 5, 6, 7, 8, 9, 10, 11, 12).

By analyzing all the combinations of 2 x 2 subtables for each 2 x 2 x d (d being the set of integers from 2 to 4), the stability of the chi-square statistic will be measured for variations in the sample sizes of the subtables. One of the items stated in the rationale for this study was that the selection of conditions and limitations of the study would be based on the extent to which the situations might actually occur in research. Since the smaller size contingency tables are found more often in the research literature, it was decided to limit the stability study to them. The larger contingency tables will be examined for the following set of 2 x 2 subtable sample sizes (12, 16, 20, 24, 28, 32, 36, 40, 44, 48).

TABLE 3.1 2 x 2 x 2 Contingency Tables

12		16		20		24		28	
12	12	12	20	12	28	12	36	12	44
		16	16	16	24	16	32	16	40
				20	20	20	28	20	36
						24	24	24	32
								28	28

32		36		40		44		48	
16	48	24	48	32	48	40	48	48	48
20	44	28	44	36	44	44	44		
24	40	32	40	40	40				
28	36	36	36						
32	32								

TABLE 3.2 2 x 2 x 3 Contingency Tables

12		
12	12	12

16		
12	16	20
16	16	16
12	12	24

20		
12	20	28
16	20	24
20	20	20
12	12	36
12	16	32
12	24	24

24		
12	24	36
16	24	32
20	24	28
24	24	24
12	12	48
12	16	44
12	20	40
12	28	32
16	16	40
16	20	36
16	28	28
20	20	32

28		
12	28	44
16	28	40
20	28	36
24	28	32
28	28	28
12	24	48
12	32	40
12	36	36
16	32	36
16	44	24
16	48	20
20	20	44
20	24	40
20	32	32
24	24	36

32		
16	32	48
20	32	44
24	32	40
28	32	36
32	32	32
12	36	48
16	36	44
16	40	40
20	28	48
20	36	40
24	24	48
24	28	44
24	36	36
28	28	40

36		
24	36	48
28	36	44
32	36	40
36	36	36
24	40	44
28	32	48
32	32	44
20	44	44
20	40	48
16	44	48
12	48	48

40		
32	40	48
36	40	44
40	40	40
36	36	48
32	44	44
24	48	48

44		
40	44	48
44	44	44
36	48	48

48		
48	48	48

TABLE 3.3 2 x 2 x 4 Contingency Tables

12			
12	12	12	12

16			
12	16	16	20
16	16	16	16
12	12	20	20
12	12	12	28

20			
12	20	20	28
16	20	20	24
20	20	20	20
12	16	20	32
12	12	20	36
16	16	20	28
12	12	12	44
16	16	16	32
12	16	16	36
12	12	16	40
12	12	24	32
12	12	28	28
16	16	24	24
12	16	24	48
12	24	20	24

24			
12	12	24	48
12	12	28	44
12	12	32	40
12	12	36	36
16	16	16	48
16	16	20	44
16	16	24	40
16	16	28	36
16	16	32	32
20	20	12	44
20	20	16	40
20	20	20	36
20	20	24	32
20	20	28	28
20	24	24	28
24	24	12	36
24	24	16	32
24	24	20	28
24	24	24	24
12	16	20	48
12	16	24	44
12	16	28	40
12	16	32	36
12	20	24	40
12	20	28	36
12	20	32	32
12	24	28	32
12	28	28	28
16	20	24	36
16	20	28	32
16	24	28	28
16	32	32	32

28			
12	12	40	48
12	12	44	44
16	16	32	48
16	16	36	44
16	16	40	40
20	20	24	48
20	20	28	44
20	20	32	40
20	20	36	36
24	24	16	48
24	24	20	44
24	24	24	40
24	24	28	36
24	24	32	32
28	28	12	44
28	28	16	40
28	28	20	36
28	28	24	32
28	28	28	28
12	16	36	48
12	16	40	44
12	20	32	48
12	20	36	44
12	20	40	40
12	24	28	48
12	24	32	44
12	24	36	40
12	28	32	40
12	28	36	36
12	32	32	36
16	20	28	48
16	20	36	40
16	24	24	48
16	24	28	44
16	24	32	40
16	24	36	36
16	28	32	36
16	32	32	32
20	24	32	36
20	28	28	36
20	28	32	32
24	28	28	32

32			
16	16	24	24
20	20	40	48
20	20	44	44
24	24	32	48
24	24	36	44
24	24	40	40
28	28	24	48
28	28	28	44
28	28	32	40
28	28	36	36
16	20	44	48
16	24	40	48
16	24	44	44
16	28	36	48
16	28	40	44
16	32	32	48
16	32	36	44
16	32	40	40
16	36	36	40
12	20	48	48
12	24	44	48
12	28	40	48
12	28	44	44
12	32	36	48
12	32	40	44
12	36	32	48
12	36	36	44
12	36	40	40
20	24	36	48
20	24	40	44
20	28	32	48
20	28	40	40
20	32	36	40
20	36	36	36
24	28	32	44
24	28	36	40
24	32	36	36
28	32	32	36
32	32	32	32

TABLE 3.3 2 x 2 x 4 Contingency Tables (continued)

36			
12	36	47	47
16	32	48	48
16	36	44	48
20	28	48	48
20	32	44	48
20	36	40	48
20	36	44	44
24	24	48	48
24	28	44	48
24	32	40	48
24	32	44	44
24	36	36	48
24	36	40	44
28	28	40	48
28	28	44	44
28	32	36	48
28	32	40	44
28	36	36	44
28	36	40	40
32	32	32	48
32	32	36	44
32	32	40	40
36	36	36	36
12	44	44	44
12	40	48	44
24	40	40	40

40			
16	48	48	48
20	44	48	48
24	40	48	48
28	36	48	48
32	32	48	48
24	44	44	48
28	40	44	48
28	44	44	44
32	36	44	48
32	40	40	48
32	40	44	44
32	44	36	48
36	36	36	48
32	36	40	44
40	40	40	40

44			
32	48	48	48
36	44	48	48
40	40	48	48
40	44	44	48
44	44	44	44

48			
48	48	48	48

TABLE 3.4 2 x 2 x 5 Contingency Tables

12	12	12	12	12
16	16	16	16	16
20	20	20	20	20
24	24	24	24	24
28	28	28	28	28
32	32	32	32	32
36	36	36	36	36
40	40	40	40	40
44	44	44	44	44
48	48	48	48	48

TABLE 3.5 2 x 2 x 6 Contingency Tables

12	12	12	12	12	12
16	16	16	16	16	16
20	20	20	20	20	20
24	24	24	24	24	24
28	28	28	28	28	28
32	32	32	32	32	32
36	36	36	36	36	36
40	40	40	40	40	40
44	44	44	44	44	44
48	48	48	48	48	48

TABLE 3.6 2 x 2 x 7 Contingency Tables

12	12	12	12	12	12	12
16	16	16	16	16	16	16
20	20	20	20	20	20	20
24	24	24	24	24	24	24
28	28	28	28	28	28	28
32	32	32	32	32	32	32
36	36	36	36	36	36	36
40	40	40	40	40	40	40
44	44	44	44	44	44	44
48	48	48	48	48	48	48

TABLE 3.7 2 x 2 x 8 Contingency Tables

12	12	12	12	12	12	12	12
16	16	16	16	16	16	16	16
20	20	20	20	20	20	20	20
24	24	24	24	24	24	24	24
28	28	28	28	28	28	28	28
32	32	32	32	32	32	32	32
36	36	36	36	36	36	36	36
40	40	40	40	40	40	40	40
44	44	44	44	44	44	44	44
48	48	48	48	48	48	48	48

TABLE 3.8 2 x 2 x 9 Contingency Tables

12	12	12	12	12	12	12	12	12
16	16	16	16	16	16	16	16	16
20	20	20	20	20	20	20	20	20
24	24	24	24	24	24	24	24	24
28	28	28	28	28	28	28	28	28
32	32	32	32	32	32	32	32	32
36	36	36	36	36	36	36	36	36
40	40	40	40	40	40	40	40	40
44	44	44	44	44	44	44	44	44
48	48	48	48	48	48	48	48	48

TABLE 3.9 2 x 2 x 10 Contingency Tables

12	12	12	12	12	12	12	12	12	12
16	16	16	16	16	16	16	16	16	16
20	20	20	20	20	20	20	20	20	20
24	24	24	24	24	24	24	24	24	24
28	28	28	28	28	28	28	28	28	28
32	32	32	32	32	32	32	32	32	32
36	36	36	36	36	36	36	36	36	36
40	40	40	40	40	40	40	40	40	40
44	44	44	44	44	44	44	44	44	44
48	48	48	48	48	48	48	48	48	48

TABLE 3.10 2 x 2 x11 Contingency Tables

12	12	12	12	12	12	12	12	12	12	12
16	16	16	16	16	16	16	16	16	16	16
20	20	20	20	20	20	20	20	20	20	20
24	24	24	24	24	24	24	24	24	24	24
28	28	28	28	28	28	28	28	28	28	28
32	32	32	32	32	32	32	32	32	32	32
36	36	36	36	36	36	36	36	36	36	36
40	40	40	40	40	40	40	40	40	40	40
44	44	44	44	44	44	44	44	44	44	44
48	48	48	48	48	48	48	48	48	48	48

TABLE 3.11 2 x 2 x 12 Contingency Tables

12	12	12	12	12	12	12	12	12	12	12
16	16	16	16	16	16	16	16	16	16	16
20	20	20	20	20	20	20	20	20	20	20
24	24	24	24	24	24	24	24	24	24	24
28	28	28	28	28	28	28	28	28	28	28
32	32	32	32	32	32	32	32	32	32	32
36	36	36	36	36	36	36	36	36	36	36
40	40	40	40	40	40	40	40	40	40	40
44	44	44	44	44	44	44	44	44	44	44
48	48	48	48	48	48	48	48	48	48	48

First Order Interaction

If the elements of a population can be classified according to three criteria A, B, C with classification A_i (i = 1, 2), B_j(j = 1, 2), and C_k (k = 1, 2, ..., d), a sample of N individuals may be classified in a three way 2 x 2 x d contingency table. If p_{ijk} represents the probabilities associated with the individual cells, and n_{ijk} is the numbers of sample elements in the individual cells; then to test whether all three criteria are mutually independent and hence no first order interaction is achieved by setting up the null hypothesis:

$$H_o: p_{ijk} = p_{i..} p_{.j.} p_{..k}$$

and testing (against the general alternative $H_A = H_o$) by the formula:

$$C^2 = \sum_{ijk} \left(n_{ijk} - \frac{n_{i..}\ n_{.j.}\ n_{..k}}{N^2}\right)^2 \Big/ \left(\frac{n_{i..}\ n_{.j.}\ n_{..k}}{N^2}\right)$$

This calculated value of chi-square is then compared against the theoretical values of chi-square with 3d-2 degrees of freedom in order to determine the level of significance for each contingency table generated.

There are 3d-2 degrees of freedom since

$$df = r c d - r - c - d + 2, \text{ and } r = c = 2.$$

Second Order Interaction

The rationale of Bartlett's test is best explained by reference to tables of size 2 x 2 x 2. In this limiting three-way case, the cross-product ratio is used as the measure of first-order interaction in each layer and the ratio of these ratios us used as the measure of second order interaction.

Thus if a 2 x 2 x 2 table has one layer with cell frequencies

n_{111}	n_{121}
n_{211}	n_{221}

and the other layer with frequencies

n_{112}	n_{122}
n_{212}	n_{222}

the first order interactions are $(n_{111}n_{221}) / (n_{121}n_{211})$ and $(n_{112}n_{222}) / (n_{122}n_{212})$ respectively, and the second order interaction is the ratio $\left[(n_{111}n_{221}) / (n_{121}n_{211})\right] / \left[(n_{112}n_{222}) / (n_{122}n_{212})\right]$ which may be written $(n_{111}n_{221}n_{122}n_{212}) / (n_{121}n_{211}n_{112}n_{222})$. Furthermore, if the interactions in the two 2 x 2 layers are identical, then $(n_{111}n_{221}) / (n_{121}n_{211}) = (n_{112}n_{222}) / (n_{122}n_{212})$.

Hence $(n_{111}n_{221}n_{122}n_{212}) \Big/ (n_{121}n_{211}n_{112}n_{222}) = 1$,

consequently $n_{111}n_{221}n_{122}n_{212} = n_{121}n_{211}n_{112}n_{222}$ and this corresponds with Bartlett's statement of the condition of zero second order interaction for the 2 x 2 x 2 case.

It is therefore required to find the minimum deviation x common to all eight cells, which will bring about this relationship for second order interaction. The value of x is accordingly determined by the equation:

$$(n_{111}-x)\ (n_{221}-x)\ (n_{122}-x)\ (n_{212}-x) \ = \ (n_{121}+x)\ (n_{211}+x)\ (n_{112}+x)\ (n_{222}+x).$$

This yields a cubic equation (with one acceptable root) which is solvable by standard numerical methods. The eight expected cell frequencies $(n_{111}-x)$, $(n_{221}-x)$, ... $(n_{222}+x)$, are now computed, whence the chi-square statistic (C^2) is:

$$C^2 = (x)^2 \left[\frac{1}{n_{111}-x} + \frac{1}{n_{121}+x} + \dots + \frac{1}{n_{222}+x}\right] \qquad (1)$$

with one degree of freedom.

Turning to the case of the table of the form 2 x 2 x 3, there are four additional observed frequencies, n_{113}, n_{123}, n_{213} and n_{223}, another degree of freedom and an additional deviation y. This will result in the following pair of simultaneous cubic equations in x and y.

$$\frac{(n_{111}-x)\ (n_{221}-x)}{(n_{121}+x)\ (n_{211}+x)} = \frac{(n_{113}+x+y)\ (n_{223}+x+y)}{(n_{123}-x-y)\ (n_{213}-x-y)} \tag{2}$$

$$\frac{(n_{112}-y)\ (n_{222}-y)}{(n_{122}+y)\ (n_{212}+y)} = \frac{(n_{113}+x+y)\ (n_{223}+x+y)}{(n_{123}-x-y)\ (n_{213}-x-y)} \tag{3}$$

Now by letting $-z = x + y$ the above pair of simultaneous equations may be written in the following form:

$$\frac{(n_{111}-x)(n_{221}-x)}{(n_{121}+x)(n_{211}+x)} = \frac{(n_{112}-y)(n_{222}-y)}{(n_{122}+y)(n_{212}+y)} = \frac{(n_{113}-z)(n_{223}-z)}{(n_{123}+z)(n_{213}+z)} \tag{4}$$

The above is true with the necessary constraint that $x + y + z = o$. To generalize the equation for the constraint it is convenient to denote the deviation by x_k. Then the pair of simultaneous cubic equations for a 2 x 2 x 3 contingency table may be represented by:

$$\frac{(n_{111}-x_1)(n_{221}-x_1)}{(n_{121}+x_1)(n_{211}+x_1)} = \frac{(n_{112}-x_2)(n_{222}-x_2)}{(n_{122}+x_2)(n_{212}+x_2)} = \frac{(n_{113}-x_3)(n_{223}-x_3)}{(n_{123}+x_3)(n_{213}+x_3)}$$

with the constraint $\sum_{k=1}^{3} x_k = 0$.

Then the ratio of the kth level of a 2 x 2 x d contingency table may be represented by:

$$r_k = \frac{(n_{11k}-x_k)\ (n_{22k}-x_k)}{(n_{12k}+x_k)\ (n_{21k}+x_k)}\ , \tag{5}$$

and $$r_k = \frac{(n_{11k})\ (n_{22k})\ (1-x_k/n_{11k})\ (1-x_k/n_{22k})}{(n_{12k})\ (n_{21k})\ (1+x_k/n_{12k})\ (1+x_k/n_{21k})} \tag{6}$$

If the binomial theorem is used on the last two factors and terms higher than the first order ignored, then:

$$r_k = \frac{(n_{11k})(n_{22k})}{(n_{12k})(n_{21k})}\left(1 - \frac{x_k}{n_{11k}} - \frac{x_k}{n_{12k}} - \frac{x_k}{n_{21k}} - \frac{x_k}{n_{22k}}\right). \tag{7}$$

Letting $$\frac{1}{g_k} = \frac{(n_{11k})\ (n_{22k})}{(n_{12k})\ (n_{21k})} \tag{8}$$

and $$\frac{1}{s_k} = \frac{1}{n_{11k}} + \frac{1}{n_{12k}} + \frac{1}{n_{21k}} + \frac{1}{n_{22k}} \tag{9}$$

then $$r_k = \frac{1}{g_k}\left(1 - \frac{x_k}{s_k}\right) \tag{10}$$

Equating approximate values of r for two levels (k and a) and solving for x_k ,

$$x_k = s_k\left[1 - \frac{g_k}{g_a}\left(1 - \frac{x_a}{s_a}\right)\right] \tag{11}$$

$$x_k = s_k - \frac{s_k g_k}{g_a} + \frac{s_k g_k x_a}{g_a s_a} \tag{12}$$

Since it is required that $\sum_k x_k = 0$,

$$0 = g_a \sum_k s_k - \sum_k (s_k g_k) + \frac{x_a}{s_a} \underset{k}{} (s_k g_k) . \tag{13}$$

Solving for x_a,

$$x_a = s_a \left[1 - \left(\frac{\sum_k s_k}{\sum_k (s_k g_k)} \right) g_a \right] \tag{14}$$

Let $h = \dfrac{\sum_k s_k}{\sum_k s_k g_k}$

Then $x_a = s_a \ (1 - hg_a)$ (15)

If none of the cells have zero frequencies in the contingency table the above equation provides approximate values of the deviation x_a in each 2 x 2 level of a 2 x 2 x d contingency table. To obtain additional corrections, the x_a must be added (algebraically) to the corresponding observed values, and these adjusted values (which are first approximations to the expected values, on the hypothesis of homogeneity) are used to repeat the above calculations. This process is continued until the corrections do not exceed 0.003. This would be an accuracy ten times greater than the precision of the probability determined by the number of iterations and therefore would not appreciably affect the precision of the probability calculations.

The chi-square for second order interaction in a $2 \times 2 \times d$ contingency table has d-1 degrees of freedom and is found by multiplying x_a^2 by the sum of the reciprocals of the expected numbers in the a*th* level of the table, and summing over the d levels.[6]

---

[6] H. W. Norton, "Calculation of Chi-Square for Complex Contingency Tables", Journal of American Statistical Association, 1945, 40, p. 257.

Zero Valued Cells

Zeros in the cells of a contingency table raise questions both experimentally and mathematically. From the experimental viewpoint the question is: are the zeros due to the size of the sample and would the zero disappear if the total sample was increased or are these zeros indicative of a forbidden level as in the sense of quantum mechanics. If a person should toss a die four times and not obtain a two, this would be an example of a zero that would disappear if the sample size was increased. However, no matter how many times a person tossed an ordinary die it would never come up with seven dots on it. Seven would be a forbidden level. The experimentalist may raise this question but if he can definitely answer it, there is no need for the experiment. From a mathematical point of view the problem is different.

In the process of division, zero produces trichotomus results depending on its rule. If the dividend is zero and the divisor is not zero, the quotient is zero. If the dividend is not zero and the divisor is zero, the quotient does not exist. If the dividend is zero and the divisor is zero, the quotient is said to be indeterminate since it is not unique.

When the expected values for second order interaction are calculated and zeros exist in the cells of the contingency table any of the above results is possible in calculating p_i. Since the study is concerned with small samples, zeros in the cells will be quite prevalent. Values for s_a and h will not exist. Norton offered a method for calculating the expected values when other "levels" (2 x 2 subtables) have a relatively large value for p_a. These problems do not exist in the calculation of the expected values for first order interaction.

In order to determine the chi-square value for second order interaction in contingency tables with small samples, a different approach than that taken by Norton must be found, since he was dealing with contingency tables that had relatively large samples in some of the levels (2 x 2 subtables).

Since in this study chi-square for first order interaction will be calculated, the second order interaction chi-square may be estimated from the first order interaction in the following manner.

The Pearson product moment correlation coefficient for all the contingency tables in a particular run which have no zero value cells will be calculated by means of:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left(N\sum X^2 - (\sum X)^2\right)\left(N\sum Y^2 - (\sum Y)^2\right)}}$$

where x is the chi-square value of first order interaction y is the chi-square value of second order interaction, then the standard deviation of the first order interactions by means of

$$s_x = \frac{1}{N}\sqrt{N\sum X^2 - (\sum X)^2} .$$

In a similar manner the standard deviation of the second order interaction chi-square

$$s = \frac{1}{N}\sqrt{N\sum Y^2 - (\sum Y)^2} .$$

This means:

$$\bar{X} = \frac{\sum X}{N} , \quad \bar{Y} = \frac{\sum Y}{N} .$$

Then the value for the second order interaction chi-square may be estimated from the first order interaction by the following:

$$Y' = r_{xy} \frac{s_y}{s_x} (X-\bar{X}) + Y$$

Then if Y' is greater than the theoretical value of chi-square for a desired probability, that particular contingency table will be so designated.

In the case of zero values and second order interaction, the coefficient of determination (r^2) is used as a measure of accuracy of the estimate of the chi-square statistic of second order interaction and the probability statement has to be modified as follows:

$$P\left(\frac{y - p}{\sqrt{\frac{P\ (1-p)}{V-V_o+r^2V_o}}} \geq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha ,$$

where V_o is the number of tables with zero valued cells.

Since in this study V is taken as 400, and accuracy of P is within 0.03, then V can be written as follows in terms of r:

$$V_o = \frac{15.84}{1-r^2}$$

This equation will be used to determine the maximum number of tables with zero valued cells allowable for this accuracy. For instance, if r = 0.5, then V_o would be 21 and if the contingency table configuration contained more than 21 tables with zero valued cells, it would be eliminated from further consideration in the calculation of probabilities for second order interaction.

Real Data

In order to show how the results of this study may be used to aid educational research, an example using real data is included in this dissertation. The data from the Loyola Correspondence Study Division chosen to be examined met the criteria of 1) availability and 2) applicability to the model. The research design was not intended solely for the examination of this data but rather the data was chosen to show how the results of this study may be used to examine this data and other qualitative data in which two of the variables may be dichotomized and the third variable having up to twelve categories.

Research in general has shown that there is little significant difference between achievement by persons who have taken courses by Correspondence Study Method, and persons who have taken courses by the more traditional methods[7,8,9.10]. While a given research study may report that for a specific course, one method is superior to the other[11]; still, correspondence regarded as a method of study and instruction contrasted with the traditional classroom and lecture technique, reveals no significant differences in measured student achievement.

Further research on the effectiveness of the correspondence method compared to other methods, then, is repetitious. However, no studies have been found concerning the evaluation of correspondence study as a self contained system answering such questions as to why certain courses are more popular; why enrollments in given courses are higher; how student motivation may relate to the completion rates for various courses, and, so forth.

---

[7] Opinion expressed by G. B. Childs in an address ("Correspondence Study: Concepts and Comments") at the Annual Meeting of the National University Extension Association, April 13, 1973.

[8] E. L. Larson, "Comparative Quality of Work Done by Students in Residence and in Correspondence Work", Journal of Educational Research, 1932, 25, pp. 105-109.

[9] T. S. Parsons, "A Comparison of Instruction by Kinescope, Correspondence Study and Customary Classroom Procedures", Journal of Educational Psychology, 1957, 48, pp. 27-40.

[10] M. Nachman and S. Opochinsky, "The Effects of Different Teaching Methods: A Methodological Study", Journal of Educational Psychology, 1958, 49, pp. 245-249.

[11] "G. B. Childs, "An Analysis of the Success in Initial University Mathematics Courses of Students With Correspondence Study and Non-Correspondence Study Backgrounds in High School," Journal of Educational Research, 1956, 49, p. 607.

The following research design is intended to show how such questions can be further clarified, if not answered. Thus, the research design serves as a method for correspondence study and provides a method of self-analysis for its own self-evaluation purposes.

One question that is of interest in this context is the following: is there any interaction between academic subjects, sex of the student and whether or not the student lives in a Catholic religious community (e.g., priest, brother or sister). The courses are grouped under four department headings used at Loyola University, Chicago. The four departments chosen are: English, Language, History and Education. These departments are chosen because the properties of their courses make them distinctive from each other. English is a subject matter that is required of all students and is classified as an academic subject. Language is an academis subject that is an elective and its product is skill building rather than content oriented. History is an elective academic subject whose product is content oriented. However, Education is a professional subject and an elective. A single year is chosen in order to hold other variables constant. The year chosen was 1961 because it is one of the large enrollment years and the data is readily available. All the students that year have student numbers that run from 7547 to 8453. These numbers are issued successively as they signed up for their first course.

A table[12] of random numbers was used to choose the last three digits of the student number. Then the name of the student may be found in an index that lists the student numbers in numerical order followed by the name of the student. Once the student's name has been found, the data concerning the courses taken by that student may be found in the master file which is a card file with the student's last name used for filing. The manner of filing Catholic religious is a little obscure, but the names may be found with a little patience and fortitude. The data will be collected and placed in a 2 x 2 x 4 contingency table and analyzed for first and second order interaction.

---

[12] M. W. Tate, Statistics in Education and Psychology, (New York: Macmillan Co., 1965), pp. 328-329.

CHAPTER 4

Results of The Study

Introduction

This chapter presents the results of the study. These results were derived from approximately 300,000 lines of computer output. From the various articles that were reviewed in Chapter Two, it is believed that the results would have been in the form of corrections for small samples to be added or subtracted to the calculated chi-square. However, this is not the case. The chi-square tests are very robust and any error that might exist is less than the experimental error of the study design for the samples considered.

In addition to these results, this study produced a computational method involving a shift of the cell frequencies that is isometric as far as the probability of interaction is concerned. Hence, the computation of the Norton routine was simplified and extended its usefulness to small samples when the iteration technique was divergent. This shift is not 100% effective, but it did work in over 99% of the cases where it was needed in this study.

The results of this study were then applied to actual data obtained from the Correspondence School of Loyola University. These results showed the usefulness of this study in cases where obtaining a small sample is laborious and a large sample by the standards of Lewis and Burke[1] may be impossible.

---

[1]D. Lewis and C. J. Burke,"The Use and Misuse of the Chi-Square Test", Psychological Bulletin, 1949, 46, p. 436.

Negative-valued Chi-Square

When the 2 x 2 x 3 contingency tables with each sub table sample size being sixteen is run on the computer, it is noted that two of the four hundred tables generated have negative valued chi-squares for second order interaction.

These negative valued chi-squares arise because the process of passing from equation (6) to equation (7) in Chapter 3 necessitates the use of the binomial theorem on the factors for the denominator of equation (6). The use of the binomial theorem requires that x_k/n_{12k} and x_k/n_{21k} be less than one for equation (6) to be a perfectly valid equivalent equation of equation (7). However, Norton's[2] iteration technique has the power of correcting for small violations of the binomial theorem. When the expected value is approximately twice as large[3] as the observed value then the iteration process has a tendency to diverge instead of converging. The large negative chi-square is due to the truncating of large numbers when division is done by Fortran IV.

---

[2]H. W. Norton, "Calculation of Chi-square for Complex Contingency Tables", Journal of American Statistical Association, 1945, 40, pp. 251-258.

[3]Ibid, p. 255.

Norton gives a method for making a correction when one of the levels for small frequencies and other levels have large frequencies. This method cannot be used here since all the levels have small frequencies. Another method was found to alleviate this problem and it depends on the symmetry of the 2 x 2 subtables. For instance, note the following 2 x 2 contingency tables:

	boys	girls
over 6 feet	10	4
under 6 feet	3	7

	boys	girls
over 6 feet	3	7
under 6 feet	10	4

The calculated chi-square would be the same for both tables. The interaction is not dependent on the labels of the columns or rows. It is found that if a problem arises from the condition of the binomial theorem being violated that it could be corrected by reversing all the rows in each subtable.

It is common practice to write 2 x 2 x n contingency tables by giving each 2 x 2 subtable a line as follows in the 2 x 2 x 3 contingency tables, where 16 is the sample size for each 2 x 2 subtable.

3	6	6	1	
5	4	6	1	(Table #1)
9	3	3	1	

If the first subtable (3 6 6 1) is displayed as a 2 x 2 contingency table it would be as follows:

3	6
6	1

The second order chi-square for table #1 is -30.55. Now, if the table is rewritten as follows:

6	3	1	6
4	5	1	6
3	9	1	3

the second order chi-square has a value of 2.11. This shift will be referred to as the "Con Midhe[4] shift".

[4]Gaelic: Hound of Meath or Strength of the Middle.

This Con Midhe shift is a very robust technique. In this study there were 1327 tables with negative valued chi-squares for second order interaction. All but four of these tables yield to the Con Midhe shift. If a table has a positive valued chi-square for second order interaction the Con Midhe shift produces either the same value or a negative value. The four tables that did not yield to the Con Midhe shift are as follows:

2 x 2 x 10 all sub-table samples equal 32

6	10	8	8
7	8	7	10
9	7	4	12
5	9	15	3
9	6	6	11
6	10	7	9
6	9	8	9
9	5	3	15
6	11	9	6
10	8	7	7

(Table #2)

2 x 2 x 12 all subtable samples equal 24

4	7	7	6
6	6	5	7
6	6	8	4
5	4	7	8
7	8	4	5
11	1	9	4
5	6	7	6
6	5	6	7
5	4	10	5
5	8	4	7
2	6	11	5
8	5	6	5

(Table #3)

2 x 2 x 12 all subtable samples equal 28

3	12	11	2
10	9	4	5
8	7	8	5
10	8	7	3
7	13	4	4
9	6	4	9
10	5	9	4
8	7	6	7
6	3	4	15
5	4	11	8
6	4	8	10
6	7	7	8

(Table #4)

2 x 2 x 12 all subtable samples equal 28

6	6	10	6
8	8	6	6
9	6	4	9
6	4	3	15
13	5	8	2
10	5	7	6
6	8	9	5
1	14	6	7
9	7	6	6
5	8	5	10
5	8	7	8
5	12	4	7

(Table #5)

There are no negative valued second order chi-squares for any of the 2 x 2 x 2 contingency tables.

The four contingency tables that are unsolvable with the Con Midhe shift are estimated by Goodman's[5] method. Goodman compared his method to Norton's[6] by calculating the chi-square for second order interaction for the same data Norton used. Goodman's value was 7.45 and Norton's value was 7.59. These were both non-significant values at the 10% level. There was no indication given in Goodman's article as to how this error would vary for small samples. However, it was assumed that calculating four values of second order chi-square by this method would not affect appreciably this study in which a total of 148,400 values of second order chi-square were calculated.

---

[5]L. A. Goodman, "Simple Methods for Analyzing Three Factor Interaction in Contingency Tables", Journal of American Statistical Association, 1964, 59, pp. 322-331.

[6]loc. cit.

The values of second order interaction chi-square for these four tables by Goodman's method are as follows:

Table	Second Order Chi-Square
#2	12.02
#3	9.05
#4	14.92
#5	11.60

None of these is significant at the 10% level. Goodman's method would never result in a negative value but no study has been done to determine a measure of its error.

Significance Levels of First and Second Order Chi-Square

In the following tables are listed the percentage of times the calculated values of the first and second order chi-squares exceeded the theoretical values of chi-square for the 10%, 5%, 2% and 1% levels of significance. Each of these percentages is based on a run of 400 tables. Therefore, each quarter of a percent is due to one table having a corresponding or larger value chi-square than the tabulated theoretical value.[7] The values that are underlined are ones that exceed the experimental error as predicted by the number of iterations chosen in the research design. There are less than 3% of the values that fall outside of the experimental error of the research design. There are some contingency tables for which the calculated percentages are not listed for second order chi-square. This is because there was a number of tables with zero valued cells that exceeded the number designated by the experimental design for the desired accuracy of this study.

---

[7]R. A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural, and Medical Research (Edinburgh: Oliver and Boyd, 1943), Table IV.

These were mostly tables with sub-table sample sizes of 12. This would yield an expected cell size of 3. Some of the tables with sub table sample sizes of 16 were also eliminated in the 2 x 2 x 3 and 2 x 2 x 4 contingency tables. For the 2 x 2 x 5 contingency tables and larger, all the tables with sub-table sample sizes of 12, 16, 20 were eliminated from the second order interaction chi-square part of the study. All these tables have an excessive number of tables with zero valued cells.

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 2

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction: Theoretical Level 10.00	5.00	2.00	1.00	Second Order Interaction: Theoretical Level 10.00	5.00	2.00	1.00
			Calculated Level							
24	12, 12	25	10.00	5.00	2.00	1.50				
32	12, 20	5.75	12.00	6.25	1.75	1.00				
32	16, 16	7	9.25	3.75	2.25	0.75	7.50	3.00	1.00	1.00
40	12, 28	11.5	9.50	3.25	0.50	0.50				
40	16, 24	3.25	10.25	3.50	1.50	0.75	11.50	4.75	1.50	0.75
40	20, 20	2.75	11.75	6.00	2.50	1.75	10.50	4.50	1.50	0.50
48	12, 36	14.5	10.25	6.50	2.25	1.25				
48	16, 32	3.5	7.75	2.50	0.75	0.25	10.25	3.50	0.50	0.00
48	20, 28	1.5	9.50	6.50	2.75	1.25	7.75	3.75	1.25	0.50
48	24, 24	1.5	9.50	3.75	2.00	1.00	13.00	6.50	2.50	1.00
56	12, 44	15	7.75	3.00	0.75	0.25				
56	16, 40	2.75	8.00	3.75	1.50	0.25	8.00	8.00	0.50	0.50
56	20, 36	1.5	12.00	5.75	2.25	1.50	13.25	6.50	2.25	1.50
56	24, 32	0	12.25	6.00	2.50	1.00	12.00	7.25	4.75	2.25
56	49, 49	0.75	13.50	7.00	3.25	1.75	13.75	6.75	4.25	1.50
64	16, 48	2.75	9.00	3.75	1.00	0.25	8.75	4.00	2.00	1.00
64	20, 44	0.5	9.75	3.75	0.50	0.25	7.75	4.25	2.25	1.75
64	24, 40	0.25	10.00	5.25	2.25	0.75	10.00	4.00	1.25	0.75
64	28, 36	0	13.00	7.50	3.75	1.25	13.75	8.00	3.75	1.25
64	32, 32	0	7.50	4.00	1.00	0.50	11.75	5.50	2.00	0.50

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 2			First Order Interaction				Second Order Interaction			
			Theoretical Level							
Table Sample Size	2 x 2 Subtable Sample Size	% Tables w/zero cells	10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
72	24, 48	0.5	8.75	4.50	2.75	1.25	9.50	4.75	1.25	0.50
72	28, 44	0	7.25	3.50	1.50	0.50	10.00	5.75	1.75	0.75
72	32, 40	0	8.75	4.50	1.75	0.50	13.50	6.25	2.00	0.75
72	36, 36	0	8.25	4.25	2.25	1.25	9.25	5.00	2.75	1.25
80	32, 48	0	8.00	5.00	2.00	1.00	12.25	6.50	3.00	1.75
80	36, 44	0	10.75	4.50	1.25	0.75	11.00	5.00	2.25	0.75
80	40, 40	0	7.25	4.00	1.25	1.00	10.00	5.00	2.00	1.00
88	40, 48	0	9.75	4.75	1.75	1.25	12.25	6.50	2.75	1.50
88	44, 44	0	11.00	6.25	2.50	0.50	11.25	7.00	3.25	2.00
96	48, 48	0	11.00	5.75	1.75	1.50	9.50	4.50	2.00	1.50

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 3

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
36	12, 12, 12	10	10.50	6.25	1.25	0.75				
48	12, 16, 20	19	9,50	5.00	1.00	0.00				
48	16, 16, 16	9.25	10.00	5.25	1.75	0.75				
48	12, 12, 24	23.75	9.75	5.75	2.50	1.00				
60	12, 20, 28	14.5	9.00	2.25	0.25	0.00				
60	12, 12, 36	25	6.75	2.75	1.75	0.50				
60	12, 16, 32	13.25	10.75	5.00	2.25	0.75				
60	12, 24, 24	13	9125	6.75	2.50	1.50				
60	16, 20, 24	8.25	11.75	5.25	1.75	1.00				
60	20, 20, 20	4.5	9.25	5.00	1.75	1.00	9.75	3.50	1.00	0.50
72	12, 24, 36	10.75	10.75	4.75	1.75	0.75				
72	12, 12, 48	24	9.25	4.50	2.25	1.00				
72	12, 16, 44	17.5	9.50	3.50	1.25	0.50				
72	12, 20, 40	14.5	10.50	6.50	2.00	1.50				
72	12, 28, 32	15.25	9.75	6.25	3.00	1.50				
72	16, 16, 40	8.25	11.75	6.00	2.25	1.25				
72	16, 24, 32	3.5	8.25	5.25	1.00	0.25	9.75	4.50	1.50	0.75
72	16, 20, 36	6	11.75	4.00	1.25	0.25	10.00	5.00	2.00	0.50
72	16, 28, 28	4.75	10.00	5.50	3.00	1.25	10.25	5.00	2.00	0.50
72	20, 20, 32	2.5	12.00	6.00	2.25	1.00	9.25	4.00	1.50	0.75
72	20, 24, 28	2	12.50	8.00	4.00	2.00	11.75	5.75	3.75	1.75
72	24, 24, 24	1.5	9.00	3.75	1.25	0.25	12.00	5.50	1.75	0.75

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 3			First Order Interaction				Second Order Interaction			
			Theoretical Level							
Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
84	12, 28, 44	12.25	9.00	3.50	1.25	0.25				
84	12, 24, 48	11.75	11.25	7.00	3.00	1.25				
84	12, 32, 40	11.25	10.75	4.25	1.00	0.50				
84	12, 36, 36	12	11.25	6.00	2.50	1.50				
84	16, 28, 40	4.5	9.25	5.00	1.25	1.00	10.00	4.00	1.50	1.00
84	16, 32, 36	3.5	7.00	3.75	1.50	0.75	9.25	5.00	1.75	0.25
84	16, 48, 20	6	8.75	2.75	1.00	0.50	7.25	2.75	0.50	0.50
84	20, 28, 36	1.25	9.00	4.25	3.00	1.50	10.00	4.00	1.25	0.75
84	20, 20, 44	1.5	10.75	5.50	1.25	0.50	11.75	5.75	2.00	1.25
84	20, 24, 40	1	10.50	5.50	2.00	1.00	13.25	5.50	3.00	1.75
84	20, 32, 32	1.25	9.75	5.25	2.50	0.25	7.00	4.00	1.50	0.50
84	24, 24, 36	0.25	10.25	4.75	1.75	0.75	10.50	4.75	1.25	0.50
84	24, 28, 32	1	9.50	3.50	1.50	0.75	14.75	6.00	1.75	0.75
84	28, 28, 28	0.75	8.50	3.25	1.50	0.75	10.75	4.25	1.75	1.25
96	12, 36, 48	10.75	9.00	3.50	1.25	0.75				
96	12, 40, 44	13.25	11.25	6.50	3.25	1.00				
96	16, 32, 48	0	9.00	4.00	1.50	0.50	10.50	3.75	0.25	0.00
96	16, 36, 44	3.75	8.25	4.25	1.00	0.00	12.00	6.25	2.25	1.25
96	20, 32, 44	0.25	10.00	4.25	1.25	0.75	7.75	3.25	1.50	1.00
96	20, 28, 48	2	10.75	5.50	1.75	0.75	9.50	3.50	1.25	0.75
96	20, 36, 40	1.5	8.50	4.00	1.50	0.75	10.75	5.00	0.75	0.00
96	24, 24, 48	0.5	8.25	4.75	1.25	0.50	10.00	5.00	2.00	1.25
96	24, 28, 44	0.75	10.00	5.25	3.00	1.25	12.25	5.75	2.75	1.75
96	24, 36, 36	0.25	11.00	5.50	1.75	1.00	10.75	5.00	1.50	0.75
96	28, 32, 36	0	10.25	4.75	1.75	1.00	10.25	5.25	2.50	1.50
96	28, 28, 40	0	9.75	5.25	2.00	1.25	9.50	4.00	2.75	1.00
96	32, 32, 32	0	10.00	5.25	1.75	1.00	12.75	7.00	2.00	0.50

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 3

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
108	12, 48, 48	14.75	11.00	5.50	1.00	0.25				
108	16, 44, 48	4.75	11.00	5.00	2.25	1.50	7.75	3.25	1.25	0.75
108	20, 40, 48	2.25	8.00	4.50	2.50	0.75	8.50	6.00	3.25	1.25
108	20, 44, 44	0.25	9.75	5.00	1.25	0.75	8.75	4.75	2.00	1.25
108	24, 36, 48	0.25	9.00	3.00	1.50	1.25	11.25	6.00	2.00	0.25
108	24, 40, 44	0.5	7.50	4.25	1.25	1.00	9.50	5.50	1.75	1.25
108	28, 32, 48	0	10.50	4.50	1.00	0.25	11.00	5.75	1.25	0.75
108	24, 40, 40	0	7.00	3.00	1.25	0.50	10.50	4.75	1.25	0.50
108	28, 36, 44	0	11.50	6.25	3.00	2.00	11.00	5.25	1.75	1.00
108	32, 36, 40	0	11.00	5.75	1.75	1.25	12.25	6.00	3.25	1.25
108	32, 32, 44	0	10.25	5.75	2.00	1.25	11.75	5.75	2.25	1.25
108	36, 36, 36	0	11.00	4.25	1.50	0.75	7.75	3.75	1.75	1.25
120	24, 48, 48	0.25	9.25	2.50	1.25	0.25	10.50	3.75	1.75	0.25
120	28, 44, 48	0	11.25	5.40	1.50	0.50	11.75	5.50	2.00	0.75
120	32, 40, 48	0	11.00	5.50	2.75	1.75	11.00	5.50	2.75	1.75
120	32, 44, 44	0	9.00	4.25	1.00	0.50	12.00	6.00	3.00	1.25
120	36, 36, 48	0	8.75	3.25	1.50	0.75	11.75	4.75	2.00	1.25
120	36, 40, 44	0	11.25	6.75	2.25	1.00	12.25	6.25	1.75	0.75
120	40, 40, 40	0	9.25	4.75	2.00	1.00	11.50	7.50	1.25	0.50
132	36, 48, 48	0	8.50	3.50	1.50	1.00	10.00	4.25	1.25	0.00
132	40, 44, 48	0	7.75	3.25	1.25	0.25	10.00	5.25	1.50	0.75
132	44, 44, 44	0	9.50	4.25	1.00	0.50	10.75	5.00	2.25	1.25
144	48, 48, 48	0	9.50	5.75	2.25	0.75	9.00	6.00	1.75	1.00

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
48	12,12,12,12	42.5	10.25	4.25	2.25	1.75				
64	12,12,12,28	32.5	8.25	3.00	1.00	0.50				
64	12,12,20,20	26.75	11.25	3.75	1.00	0.00				
64	12,16,16,20	19.25	12.00	5.00	3.75	1.25				
64	16,16,16,16	13.25	10.50	5.25	2.25	1.75				
80	12,12,12,44	33.75	10.50	4.50	1.25	0.25				
80	12,12,20,36	23.5	7.75	4.00	1.25	0.00				
80	12,12,16,40	24.5	9.75	5.50	2.75	1.50				
80	12,12,24,32	25.5	6.25	1.50	0.50	0.50				
80	12,12,28,28	23.25	9.25	5.50	2.25	1.25				
80	12,20,20,28	14.75	9.50	4.25	2.25	1.00				
80	12,16,20,32	15.0	10.25	6.25	2.25	1.00				
80	12,16,16,36	22.0	12.75	5.75	1.00	0.25				
80	12,16,24,28	16.25	12.50	6.25	4.25	1.50				
80	12,24,20,24	16.25	11.25	5.25	3.50	2.25				
80	16,16,16,32	13.25	10.50	6.50	2.75	1.25				
80	16,16,24,24	10.0	11.00	5.25	1.50	0.75				
80	16,16,20,28	9.75	9.00	4.50	0.25	0.00				
80	16,20,20,24	7.5	9.75	3.50	1.50	0.50				
80	20,20,20,20	4.75	9.50	5.50	2.00	0.50	11.25	4.50	2.00	1.00
96	12,12,24,48	26.25	13.25	7.25	2.25	1.00				
96	12,12,28,44	24.0	9.00	5.75	3.00	0.75				
96	12,12,32,40	26.0	9.50	2.50	1.25	0.75				
96	12,12,36,36	29.0	10.75	6.25	2.50	1.50				

Theoretical vs. Calculated Levels of Significance

Table Dimension
2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
96	12,20,20,44	15.25	10.25	4.50	2.50	2.00				
96	12,24,24,36	13.0	10.75	4.50	1.00	0.50				
96	12,16,20,48	18.75	11.00	3.75	1.50	0.75				
96	12,16,24,44	16.75	7.25	3.00	0.75	0.00				
96	12,16,28,40	17.0	10.75	4.00	1.25	0.75				
96	12,16,32,36	18.0	7.75	4.50	1.50	0.75				
96	12,20,24,40	14.75	9.25	4.75	2.00	1.25				
96	12,20,28,36	11.0	7.25	1.75	1.00	0.50				
96	12,20,32,32	15.25	9.25	5.00	2.50	1.25				
96	12,24,28,32	7.5	8.50	2.25	0.25	0.00				
96	12,28,28,28	13.25	6.75	3.25	1.00	0.25				
96	16,16,16,48	9.0	9.25	4.00	1.75	0.25				
96	16,16,20,44	6.0	9.25	4.75	1.50	0.75				
96	16,16,24,40	6.75	9.00	5.00	2.25	0.75				
96	16,16,32,32	7.5	7.50	2.75	0.50	0.50				
96	16,20,20,40	8.75	9.75	5.25	1.75	0.75				
96	16,16,28,36	5.5	9.00	4.50	1.75	0.75	7.50	3.25	2.00	1.00
96	16,24,24,32	3.25	9.25	4.75	2.00	1.50	10.50	4.50	2.25	0.25
96	16,20,24,36	6.25	10.75	6.00	2.25	1.75	10.75	6.00	2.25	1.75
96	16,20,28,32	5.5	10.00	4.25	1.75	1.25	7.50	4.00	1.50	0.75
96	16,24,28,28	5.0	10.00	4.50	1.25	0.75	9.50	3.25	1.00	0.75
96	16,32,32,32	4.75	8.75	5.25	2.75	0.50	7.50	3.50	1.00	0.25
96	20,20,20,36	3.75	8.00	3.75	1.50	0.00	11.50	4.75	2.25	1.00
96	20,20,24,32	3.5	12.50	6.25	2.25	1.25	11.00	4.50	2.00	1.25
96	20,20,28,28	3.75	9.25	5.00	2.25	1.25	8.75	5.00	2.00	0.75
96	20,24,24,28	2.0	11.25	6.25	1.25	0.75	8.25	4.25	1.50	1.00
96	24,24,24,24	1.0	8.00	3.75	1.75	1.00	11.75	5.75	2.25	0.50

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction 10.00	5.00	2.00	1.00	Second Order Interaction 10.00	5.00	2.00	1.00
			Calculated Level							
112	12,12,40,48	23.5	8.75	2.75	0.50	0.25				
112	12,12,44,44	22.0	8.50	4.50	2.25	0.75				
112	12,28,28,44	13.25	9.50	4.75	2.00	1.00				
112	12,16,36,48	16.0	10.50	5.50	2.00	1.25				
112	12,16,40,44	14.5	9.00	4.75	1.25	0.00				
112	12,20,32,48	16.5	8.75	5.00	2.25	1.25				
112	12,20,36,44	13.0	9.75	3.75	1.50	0.75				
112	12,20,40,40	14.0	8.00	3.25	1.75	0.75				
112	12,24,28,48	13.0	8.50	4.00	1.00	0.25				
112	12,24,32,44	17.5	10.50	5.00	1.75	1.00				
112	12,24,36,40	15.0	10.25	5.75	3.25	2.50				
112	12,28,32,40	14.0	11.50	5.25	2.25	0.50				
112	12,28,36,36	13.75	8.75	3.50	0.75	0.00				
112	12,32,32,36	13.0	10.50	5.50	2.00	1.00				
112	16,16,32,48	8.25	11.50	5.50	1.50	1.00				
112	16,16,36,44	9.5	10.50	5.50	2.50	1.25				
112	16,16,40,40	6.5	9.25	3.75	0.50	0.00				
112	16,20,28,48	6.25	10.25	3.50	1.50	0.25				
112	16,20,32,44	6.5	11.00	5.25	1.25	0.25				
112	20,20,24,48	1.5	10.50	3.50	0.50	0.25	11.50	6.00	3.00	0.75
112	20,20,28,44	3.0	8.75	3.00	0.50	0.25	8.50	3.50	1.25	1.00
112	20,20,32,40	1.25	7.75	2.50	1.50	0.75	10.25	5.00	1.75	0.75
112	20,20,36,36	1.75	10.25	4.50	1.75	1.00	10.00	6.00	2.00	0.25
112	24,24,16,48	5.5	7.75	4.50	1.75	0.50	11.00	4.25	1.25	0.50
112	24,24,20,44	2.25	9.25	4.25	2.00	1.25	11.50	5.75	2.50	1.00
112	24,24,24,40	2.0	8.50	3.50	1.50	0.75	7.25	2.75	0.25	0.25
112	24,24,28,36	0.75	10.50	5.25	1.75	1.25	10.25	4.00	2.50	1.50
112	24,24,32,32	0.75	9.75	4.75	2.25	1.25	10.25	4.50	2.50	0.75
112	28,28,20,36	1.25	10.75	4.50	1.00	0.25	8.00	3.50	1.75	0.75
112	28,28,16,40	3.5	9.00	4.00	2.00	1.25	10.00	3.75	0.75	0.75
112	28,28,24,32	0.25	10.75	5.00	2.75	1.00	9.50	5.00	1.75	0.25

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction: Theoretical Level 10.00	5.00	2.00	1.00	Second Order Interaction: 10.00	5.00	2.00	1.00
			Calculated Level							
112	16,28,28,40	4.25	13.25	7.00	3.75	1.50	8.75	4.25	1.25	1.25
112	16,28,32,36	3.0	10.25	4.50	2.00	0.50	10.25	4.75	2.25	0.75
112	16,32,32,32	2.25	8.25	3.25	1.00	0.25	11.50	5.00	1.75	0.50
112	20,24,32,36	1.5	7.75	4.25	1.50	0.75	9.00	5.50	1.75	1.00
112	20,28,28,36	2.0	8.75	4.75	1.50	0.50	12.00	5.75	1.50	0.50
112	20,28,32,32	0.75	10.25	5.25	2.50	1.50	9.75	4.50	1.00	0.00
112	28,28,28,28	0.	9.25	4.75	2.25	1.75	10.50	5.00	2.00	1.25
128	12,20,48,48	14.0	8.00	4.00	1.50	0.50				
128	12,24,44,48	13.75	8.25	3.75	1.75	0.50				
128	12,28,40,48	13.25	8.00	3.50	1.50	0.50				
128	12,28,44,44	12.0	9.50	6.00	2.75	0.75				
128	12,32,36,48	14.75	11.00	5.00	2.25	1.50				
128	12,32,40,44	13.25	10.50	4.75	1.25	0.25				
128	12,36,32,48	13.0	13.75	7.75	3.50	2.00				
128	12,36,36,44	11.5	8.00	3.25	1.25	0.50				
128	12,36,40,40	14.75	8.75	3.75	1.25	0.50				
128	20,20,40,48	1.0	10.00	4.00	2.25	1.25	11.25	5.50	1.50	0.50
128	20,20,44,44	2.25	13.00	6.00	2.25	1.25	11.50	5.25	1.50	0.50
128	24,24,32,48	1.5	8.25	4.50	2.75	1.25	11.00	4.75	2.00	1.25
128	24,24,36,44	1.0	8.00	3.75	1.50	0.75	10.50	4.00	0.75	0.00
128	24,24,40,40	0.75	10.00	6.25	2.25	1.00	12.00	6.75	2.75	1.50
128	28,28,24,48	0.5	9.75	5.75	3.00	1.00	13.25	6.50	3.25	1.25
128	28,28,28,44	0.5	9.75	5.25	2.75	2.00	13.75	6.00	2.00	1.25
128	28,28,32,40	0.	9.75	4.75	2.00	1.25	11.00	3.50	1.75	1.25
128	28,28,36,36	0.5	9.75	6.50	2.75	2.25	13.00	6.25	2.00	0.50
128	16,20,44,48	4.75	9.75	4.75	2.00	0.50	9.25	5.50	2.00	1.75

Theoretical vs. Calculated Levels of Significance

Table Dimension
2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
128	16,24,40,48	5.75	8.25	4.00	1.50	0.25	7.75	3.25	1.00	0.75
128	16,24,44,44	5.5	7.00	3.00	1.25	0.50	10.25	5.75	2.50	0.75
128	16,28,36,48	2.0	8.50	4.50	0.75	0.00	8.00	3.75	1.75	1.00
128	16,28,40,44	5.25	7.25	4.00	1.00	0.25	9.50	3.75	0.75	0.25
128	16,32,32,49	4.25	8.25	3.50	1.50	0.75	10.00	3.25	1.25	0.50
128	16,32,36,44	2.5	11.00	5.50	2.25	0.50	8.75	5.25	1.50	0.50
128	16,32,40,40	6.0	9.50	6.00	1.75	0.75	8.50	5.00	2.00	1.25
128	16,36,36,40	3.5	9.00	4.75	2.00	0.75	9.50	5.75	2.25	0.75
128	20,24,36,48	1.0	12.00	5.25	2.25	2.00	10.25	5.75	2.25	1.25
128	20,24,40,44	1.75	10.75	4.50	1.25	0.75	10.75	4.50	1.75	1.00
128	20,28,32,48	2.0	11.50	5.25	2.25	1.75	8.75	4.25	2.75	1.75
128	20,28,36,44	1.25	10.50	5.50	2.50	1.25	8.50	4.00	1.50	1.25
128	20,28,40,40	1.25	10.00	4.00	1.50	1.00	10.75	5.50	2.50	0.75
128	20,32,32,44	1.0	8.50	4.75	2.75	1.25	10.25	4.75	2.25	1.00
128	20,32,36,40	1.5	13.50	5.50	3.00	1.75	13.50	7.00	3.25	1.00
128	20,36,36,36	1.0	8.50	5.25	2.75	1.75	8.50	4.00	1.50	1.00
128	24,28,32,44	0.5	10.75	5.00	1.00	0.75	13.25	5.25	2.50	1.25
128	24,28,36,40	0.25	10.00	5.00	1.50	0.75	11.00	6.50	2.00	1.50
128	24,32,32,40	0.25	13.50	7.50	2.50	0.75	10.25	7.00	1.00	0.25
128	24,32,36,36	0.5	10.00	7.00	2.50	1.25	10.25	5.25	2.75	0.75
128	28,32,32,36	0.25	10.25	5.00	1.75	1.25	10.75	5.50	3.25	2.00
128	32,32,32,32	0.25	11.25	5.25	2.00	0.50	11.00	4.75	2.50	1.00

Theoretical vs. Calculated Levels of Significance

Table Dimension
2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
144	12,36,48,48	11.5	8.00	4.00	0.75	0.00				
144	12,44,44,44	15.0	10.00	5.75	3.00	1.25				
144	12,40,48,44	11.25	8.75	3.50	1.25	0.75				
144	16,32,48,48	2.75	10.00	6.25	2.25	1.25	12.25	6.50	3.00	2.00
144	16,36,44,48	3.75	12.00	5.75	1.25	0.50	12.75	5.50	2.25	0.75
144	20,28,48,48	3.25	8.75	3.50	1.25	1.00	9.75	4.25	2.75	1.50
144	20,32,44,48	1.5	8.00	4.00	1.75	0.50	9.00	4.25	1.00	0.00
144	20,36,40,48	0.75	9.75	4.00	1.50	0.75	9.50	4.75	1.50	1.50
144	20,36,44,44	0.75	9.00	3.00	1.25	0.75	12.50	7.00	3.25	2.50
144	24,24,48,48	0.0	10.50	5.50	2.50	1.00	11.25	4.75	2.25	1.00
144	24,28,44,48	0.75	10.75	6.25	3.00	1.50	12.75	6.75	2.75	1.75
144	24,32,40,48	0.5	8.50	3.75	1.25	0.50	10.25	5.50	3.25	2.50
144	24,32,44,44	0.0	11.25	5.50	2.75	1.50	10.00	6.75	2.25	1.50
144	24,36,36,48	0.75	10.25	5.75	2.50	1.00	11.00	6.25	2.75	1.25
144	24,36,40,44	0.5	8.50	4.00	2.00	1.25	10.25	5.25	1.50	0.50
144	24,40,40,40	0.25	10.75	6.25	1.25	0.50	12.25	5.75	2.50	1.50
144	28,28,40,48	0.25	7.75	4.75	2.00	1.50	9.50	3.00	1.00	1.00
144	28,28,44,44	0.0	10.50	5.00	0.25	0.25	9.75	5.25	2.50	1.25
144	28,32,36,48	0.0	11.00	5.75	1.75	0.75	11.25	6.00	2.50	1.00
144	28,32,40,44	0.0	9.75	4.00	2.25	0.75	11.25	5.50	2.50	1.75
144	28,36,36,44	0.25	9.25	5.25	2.50	0.50	10.25	4.00	2.50	1.00
144	28,36,40,40	0.0	12.25	7.75	2.25	0.75	11.00	6.00	2.75	0.75
144	32,32,32,48	0.25	11.50	6.75	2.75	1.50	10.00	5.50	2.50	1.75
144	32,32,36,44	0.25	10.25	6.00	3.50	1.75	10.25	6.00	2.00	0.25
144	32,32,40,40	0.0	10.50	5.25	1.25	1.25	9.75	6.00	2.75	1.50
144	36,36,36,36	0.0	9.00	4.25	2.25	0.50	9.00	4.75	2.25	1.00

Theoretical vs. Calculated Levels of Significance

Table Dimension 2 x 2 x 4

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction 10.00	5.00	2.00	1.00	Second Order Interaction 10.00	5.00	2.00	1.00
			Calculated Level							
160	16,48,48,48	4.5	8.50	3.50	1.75	1.00	8.00	3.00	1.25	0.50
160	20,44,48,48	1.75	7.75	4.25	1.75	0.50	10.25	5.00	1.50	1.00
160	24,40,48,48	0.25	12.50	5.25	2.25	1.25	8.50	4.00	1.25	0.25
160	24,44,44,48	0.25	10.50	4.75	2.25	1.00	11.50	6.25	3.25	1.25
160	28,36,48,48	0.0	10.25	4.50	0.75	0.50	12.75	6.50	3.00	2.00
160	28,40,44,48	0.0	10.25	3.75	0.75	0.25	12.25	7.00	2.75	1.25
160	28,44,44,44	0.0	11.75	5.75	2.25	0.50	11.50	4.75	3.00	1.50
160	32,32,48,48	0.0	9.25	4.00	2.00	1.00	10.50	5.00	3.50	1.75
160	32,36,44,48	0.25	9.00	5.00	1.75	0.00	12.75	6.75	2.75	1.25
160	32,40,40,48	0.0	12.75	7.00	2.75	1.50	12.75	5.50	1.50	0.75
160	32,40,44,44	0.0	9.50	5.00	1.50	0.75	9.50	4.75	1.75	1.00
160	32,36,40,44	0.0	11.00	4.75	2.00	0.75	12.50	4.50	2.00	1.00
160	36,40,36,48	0.0	7.50	4.75	1.25	0.75	11.75	4.75	1.75	0.50
160	40,40,40,40	0.0	12.00	12.00	2.25	1.00	13.00	8.25	4.50	2.00
176	32,48,48,48	0.0	10.75	5.50	2.00	0.75	10.50	6.75	3.50	1.50
176	36,44,48,48	0.0	8.75	3.75	1.75	0.25	9.25	5.75	2.75	1.25
176	40,40,48,48	0.0	8.50	5.25	2.25	0.25	9.25	4.75	2.25	1.50
176	40,44,44,48	0.0	8.75	4.25	1.50	1.00	9.25	3.75	1.50	0.50
176	44,44,44,44	0.0	12.75	5.75	3.25	2.00	9.50	5.00	2.75	1.75
192	48,48,48,48	0.0	10.25	5.25	2.25	1.00	14.00	7.25	3.75	1.75

Theoretical vs. Calculated Levels of Significance

Table Dimension

2 x 2 x 5

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
60	12	50.5	9.50	4.00	1.25	0.50				
80	16	19.75	8.00	2.50	1.00	0.50				
100	20	7.0	9.00	3.75	1.50	1.00				
120	24	1.25	10.50	3.50	0.75	0.50	11.75	6.50	1.50	1.00
140	28	0.5	9.75	4.25	1.75	0.75	11.50	7.25	1.75	1.00
160	32	0.0	12.75	6.00	2.75	1.00	8.75	5.25	2.25	2.00
180	36	0.0	7.75	3.75	1.50	0.50	9.50	5.25	2.75	1.25
200	40	0.0	10.50	5.25	2.00	1.00	11.75	6.75	3.00	0.75
220	44	0.0	9.75	4.25	1.25	0.50	10.75	6.75	1.50	0.75
240	48	0.0	8.50	5.00	1.50	1.50	9.00	5.25	1.75	1.00

2 x 2 x 6

72	12	53.5	9.00	4.00	2.00	0.75				
96	16	23.0	9.00	3.50	2.25	1.00				
120	20	8.25	10.25	7.25	2.75	2.25				
144	24	1.0	8.50	5.50	2.25	0.25	7.25	3.25	0.25	0.00
168	28	1.0	9.50	4.25	1.75	1.25	9.25	5.00	2.50	1.50
192	32	0.0	9.00	3.50	1.25	0.50	8.50	4.50	1.75	0.50
216	36	0.25	8.75	4.75	1.50	0.75	8.75	6.75	3.00	1.50
240	40	0.0	10.25	4.75	1.75	0.50	10.25	5.75	3.25	2.00
264	44	0.0	9.75	3.25	0.75	0.50	9.75	5.50	2.25	0.50
288	48	0.0	12.00	5.25	2.75	1.00	12.00	6.00	2.50	1.50

Theoretical vs. Calculated Levels of Significance

Table Dimension

2 x 2 x 7

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction: Theoretical Level 10.00	5.00	2.00	1.00	Second Order Interaction: Theoretical Level 10.00	5.00	2.00	1.00
			Calculated Level							
84	12	64.0	9.00	5.00	2.50	1.00				
112	16	25.5	8.25	3.75	1.50	0.75				
140	20	7.75	10.75	6.00	2.25	1.00				
168	24	3.25	10.50	3.00	1.50	0.75	16.50	7.25	3.50	2.00
196	28	1.25	10.00	5.25	2.25	1.00	10.50	4.50	1.50	0.75
224	32	0.75	9.75	3.50	1.50	0.00	11.50	6.25	2.25	1.00
252	36	0.0	11.25	6.25	1.75	0.50	11.25	6.25	2.75	1.50
280	40	0.0	10.00	5.25	1.25	1.00	13.50	7.75	3.75	1.75
308	44	0.0	8.25	4.25	0.75	0.25	11.50	5.00	3.75	2.00
336	48	0.0	13.25	6.00	2.50	1.25	10.50	5.25	2.25	1.50

2 x 2 x 8

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order 10.00	5.00	2.00	1.00	Second Order 10.00	5.00	2.00	1.00
96	12	68.75	8.00	3.25	1.25	0.50				
128	16	26.0	7.00	2.50	1.00	0.75				
160	20	8.75	10.75	5.00	2.50	1.75				
192	24	3.25	11.50	5.50	1.50	0.75	9.75	5.50	1.50	0.75
224	28	1.75	10.25	5.75	3.25	1.00	11.00	6.50	1.50	0.50
256	32	0.5	11.00	6.75	2.75	0.75	13.00	6.75	2.50	1.50
288	36	0.0	11.50	6.00	1.75	0.50	9.50	5.25	2.00	0.75
320	40	0.0	11.00	5.25	2.50	1.25	11.50	5.25	1.50	0.75
352	44	0.0	11.75	5.50	2.50	1.00	10.75	6.75	3.25	1.50
384	48	0.0	7.25	3.50	0.75	0.25	9.75	4.50	1.50	1.00

Theoretical vs. Calculated Levels of Significance

Table Dimension

2 x 2 x 9

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction				Second Order Interaction			
			Theoretical Level							
			10.00	5.00	2.00	1.00	10.00	5.00	2.00	1.00
			Calculated Level							
108	12	70.75	9.50	4.75	1.25	0.25				
144	16	31.5	10.75	4.50	2.25	1.50				
180	20	11.25	9.50	4.00	2.00	1.00				
216	24	2.25	11.25	7.00	3.00	0.75	8.75	5.25	2.00	1.00
252	28	1.0	9,50	5.00	2.25	0.75	10.75	4.50	2.50	1.25
288	32	0.0	10.75	4.75	1.50	1.00	11.50	4.50	2.25	1.25
324	36	0.0	8.00	4.00	1.50	0.50	11.50	6.50	2.50	0.75
360	40	0.0	10.00	3.75	1.50	1.00	9.50	6.00	3.00	1.75
396	44	0.0	10.75	6.25	2.75	1.25	11.25	7.00	1.75	1.25
432	48	0.0	9.75	4.50	2.00	1.25	10.00	6.75	3.25	1.25

2 x 2 x 10

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order 10.00	5.00	2.00	1.00	Second Order 10.00	5.00	2.00	1.00
120	12	72.5	9.50	3.75	1.00	0.50				
160	16	23.5	7.50	3.50	1.25	0.50				
200	20	16.0	7.00	4.75	1.25	1.00				
240	24	3.5	10.00	4.75	1.50	0.25	10.00	4.50	2.00	0.25
280	28	0.75	9.25	5.00	2.25	1.00	10.75	4.75	2.75	1.25
320	32	0.75	8.00	3.25	1.25	0.50	10.50	5.25	2.50	0.50
360	36	0.5	13.25	7.50	3.75	1.75	11.25	5.50	1.75	0.75
400	40	0.0	10.50	4.00	1.00	0.25	10.75	5.50	3.00	1.00
440	44	0.0	11.50	7.00	3.75	1.75	10.25	6.75	3.50	1.00
480	48	0.0	9.00	4.25	0.75	0.25	10.00	4.50	1.25	0.25

Theoretical vs. Calculated Levels of Significance

Table Sample Size	2x2 Subtable Sample Size	% Tables w/zero cells	First Order Interaction (Theoretical Level) 10.00	5.00	2.00	1.00	Second Order Interaction (Theoretical Level) 10.00	5.00	2.00	1.00
Table Dimension 2 x 2 x 11			Calculated Level							
132	12	71.5	8.25	1.50	0.50	0.25				
176	16	34.75	8.50	4.75	1.00	0.50				
220	20	13.25	9.75	5.00	2.25	1.00=				
264	24	3.5	10.75	5.75	2.75	1.50	9.25	5.50	2.50	0.50
308	28	2.25	12.50	5.50	0.75	0.75	14.25	5.75	2.50	0.75
352	32	0.0	7.50	3.00	1.00	0.50	8.50	5.50	1.75	0.50
396	36	0.0	8.25	4.25	1.75	1.25	11.75	5.25	1.25	0.50
440	40	0.5	9.00	3.75	1.25	0.75	9.00	5.25	1.00	1.00
484	44	0.0	13.50	6.75	3.50	1.50	12.00	6.50	2.50	1.50
528	48	0.0	9.25	4.25	1.25	0.75	10.50	6.75	2.00	1.00
2 x 2 x 12										
144	12	80.75	8.00	3.50	1.50	0.75				
192	16	40.72	8.75	4.50	2.00	0.75				
240	20	21.0	9.75	5.00	2.00	0.75				
288	24	6.0	9.00	4.50	2.25	0.75	9.75	5.25	1.25	0.75
336	28	2.0	10.75	5.25	2.25	0.75	12.75	5.00	2.25	1.00
384	32	0.25	8.75	4.50	2.50	0.75	10.50	5.75	1.75	0.50
432	36	0.0	11.25	4.50	2.25	0.75	11.50	4.25	1.75	1.25
480	40	0.0	9.75	4.50	1.00	0.75	11.75	6.75	1.50	0.75
528	44	0.0	13.75	6.75	3.00	2.00	10.50	5.75	3.00	1.50
576	48	0.0	10.25	4.75	2.50	1.00	10.50	6.00	3.75	1.25

Correspondence School Study

In order to do this study data on 148 persons enrolled in 1961 were obtained from the office of the Correspondence School of Loyola University. Twenty-five man hours were required to obtain this data. Of the 148 persons, 88 persons took courses that were of interest in the study. The data were categorized as follows:

Education courses

	Religious	Lay
completed course	8	10
uncompleted course	1	5

English courses

	Religious	Lay
completed course	2	13
uncompleted course	2	7

History courses

	Religious	Lay
completed course	5	5
uncompleted course	1	5

Language courses

	Religious	Lay
completed course	8	7
uncompleted course	3	6

First order interaction chi-square is 10.45. Second order interaction chi-square is 2.50. These results indicate that there is neither first order or second order interaction in this 2 x 2 x 4 contingency table.

Until the previous Monte Carlo investigation was performed, there may have been a lack of confidence in the preceding statements because of the published assertions of Cochran[8] and Lewis and Burke[9] concerning sample sizes. However, the results of the Monte Carlo distribution in the previous section of this chapter show that, when the individual cell samples are as small as one, chi-square values of this small a magnitude indicate a lack of interaction.

---

[8]W. G. Cochran, "Some Methods for Strengthening the Common Chi-square Test", Biometrics, 1954, 10, pp. 417-451.

[9]D. Lewis and C. J. Burke, "The Use and Misuse of the Chi-square Test", Psychological Bulletin, 1949, 46, pp. 433-489.

CHAPTER 5

Conclusions and Recommendations for Future Research

Conclusions

From the results shown in Chapter Four it is evident that the chi-square test for first order interaction is quite robust as far as the sample size is concerned. When the expected value for each cell is as small as three, the results are within the experimental error of the design even when the contingency table is as large as a 2 x 2 x 12. This study indicates that for first order interaction the chi-square test is valid even when the expected value for each cell is as low as three and over 80% of the tables have zero valued cells as in the case of the 2 x 2 x 12 contingency tables with 2 x 2 subtable sample sizes of 12. There is no indication that the chi-square test for second order interaction would be any less robust if another method besides regression analysis, which was used in this study, could be used for determining second order interaction when any of the cells are zero valued. However, the results indicate that as long as all the cells have a minimum value of one the chi-square for the second order interaction is within the limits of the error expected for the number of iterations used in the study as discussed in Chapter Three in the section titled The Number of Iterations.

The homogeneity of the correspondence school data is evident from its low values for both first and second order interaction chi-square. In both cases the probability that the configuration is due to non-chance is less than 70% if the standard chi-square table is used. However, this study only concerned itself with probabilities of 10% and less. Therefore, the lack of interaction at the 10% or less level of significance is assured but nothing can be said concerning the 30% value. This is an example of what is meant by the statement that: Monte Carlo techniques do not yield results that are as general as analytic methods.

This study may also be used to determine how small a sample size may be used to determine second order interaction. The research practitioner may use the table in Chapter Four in the following manner. first, determine the dimensions of the contingency table under consideration; secondly, determine the total sample size and the corresponding sub-table sample sizes for which the second order interaction values of chi-square have been tabulated. For instance, if the second order interaction is to be investigated between three variables by means of a 2 x 2 x 4 contingency table with a total sample size of 144, the smallest subtable sample size may be 16. A subtable sample of 12 would be too small for its accuracy to have been verified by this study.

When greater accuracy is desired, for example, if the results are required to be within one-half of one percent, 99 percent of the time, the program listed in Appendix One should be run 9566 times and these results used. If a choice may be made concerning the subtable sizes, then the combination with the smallest percentage of tables with zero valued cells should be used. For example, if a study is to be done involving a 2 x 2 x 4 contingency table with a total sample size of 128, more accuracy may be obtained using 2 x 2 subtable samples of 28, 28, 32, 40 than subtable samples of 16, 20, 44, 48.

The column of the results labeled "% Tables w/Zero Cells" may be used to obtain the probability of obtaining a contingency table with a zero valued cell for a given sample size and contingency table dimensions. For instance, if a research study necessitates the use of a 2 x 2 x 4 contingency table and the data results in the following 2 x 2 subtable sample sizes - 16, 28, 36, 48, then 2% of the time such a table is constructed randomly, one or more of the cells will be zeroed valued.

The program in Appendix Two may be used to obtain the chi-square value for second order interaction. It is the same routing used in Appendix One for calculating second order interaction. The rewriting of the program in Appendix Two is for the convenience of the reader of this study.

<u>Recommendations for Future Research</u>

1. Comparison of Goodman's Method with Norton's Routine

It would be of interest to determine the error of Goodman's method for calculating the chi-square for second order interaction using Norton's method as the standard. Goodman's method has the advantages of not generating any negative valued chi-squares and it is easier to compute. The disadvantage is that the results do not equal the results of Norton's method and there is not any means for determining the possible error.

2. Consideration of j x k x d Parallelpiped Contingency Tables

Further research should be conducted in expanding this study from 2 x 2 x d contingency tables to three dimensional contingency tables where each of the variables has more than two categories. This would first necessitate a computer program that could be an extension of Kastenbaum's and Lamphiear's[1] work. Here the method of correcting for negative valued chi-squares may not be as simple as it was in the 2 x 2 x d case.

---

[1] M. A. Kastenbaum and D. E. Lamphiear, "Calculation of Chi-square to Test the No Three Factor Interaction Hypothesis ", <u>Biometrics</u>, 1959, 15, pp. 107-115.

APPENDIX I

Program For Obtaining Significance Levels Of First and Second Order Chi-Square

```
      DATA STARS,BL /'**  ',' '/
      DIMENSION NCELL(4),OBS(2,2,12),TAU(16),A(12),B(12),C(12),D(12)
      DIMENSION N(12),N1(12),SAVE(4,12)
      DIMENSION CHISQ1(400),CHISQ2(400),FLAG(400)
      DIMENSION NSAVE (20,4,12),CHITST(4),NFREQ(4),CHTST2(4),NFRQ2(4)
      IY = 65539
      NLINES = 0
      ZERO = 0.0
C READ IN NUMBER OF  2X2 LEVELS , NUMBER OF TABLES TO BE GENERATED,
C .& THE SEED VALUE FOR THE RANDOM NUMBER GENERATOR
C                          NPUNCH=0 DO NOT PUNCH CHISQUARES
C                          NPUNCH=1 PUNCH CHISQUARES
C  THIS PROGRAM WILL GENERATE A MAXIMUM OF 400 TABLES, EACH TABLE OF 12 LEVELS
C  FOR GREATER VALUES CHANGE CARDS  2,3,4,5,11,193
    3 READ (1,999,END=2000) NLEVEL,NRUNS,NPUNCH,IX
      NEGATI = 0
      IF (NRUNS.LE.400) GO TO 100
      WRITE (3,889)
      GO TO 2000
  100 CONTINUE
      DO 6 I = 1,12
    6 N(I) = 0
      DO 8 I = 1,4
      NFRQ2(I) = 0
    8 NFREQ(I) = 0
C  READ SAMPLE SIZES AT EACH 2X2 LEVEL
      READ (1,998) (N(I),I = 1,NLEVEL)
C  READ 4 SIGNIFICANT VALUES FOR EACH CHISQUARE FREQUENCY TABLE
      READ (1,997) (CHITST(I),I = 1,4),(CHTST2(I),I = 1,4)
      DO 4 I = 1,NLEVEL
      N1(I) = N(I)
      N(I) = N(I)*4
    4 CONTINUE
      WRITE (3,891)
      DO 7 I = 1,NRUNS
    7 FLAG(I) = BL
C  THIS LOOP GENERATES THE CONTINGENCY TABLES AND CALUATES THE FIRST AND
C  SECOND ORDER CHISQUARES
      DO 1100 NRUN = 1,NRUNS
      NTOT = 0
      NTST = 0
      NSUBT = 1
    2 IF (NSUBT.GT.NLEVEL) GO TO 1000
      DO 1 I = 1,5
    1 NCELL (I) = 0
    5 IF (NTOT.GE.N(NSUBT)) GO TO 20
C  CALL THE RANDOM NUMBER GENERATOR
      CALL RANDU (IX,IY,NUM)
C  DISCARD ALL VALUES GREATER THAN 4
C  FILL UP THE CELLS OF THE CONTINGENCY TABLE
      IF (NUM.GT.3) GO TO 5
```

```
   10 NCELL (NUM+1) = NCELL (NUM+1) + 1
      NTOT = NTOT + 1
      GO TO 5
   20 CONTINUE
      DO 21 I = 1,4
C  TEST FOR ZERO ELEMENTS IN THE TABLE
      IF (NCELL(I).EQ.0) NTST = 1
   21 CONTINUE
C  FILL UP THE A,B,C, AND D TABLES FOR USE BY THE SECOND ORDER CHISQUARE
C  SUBROUTINE
      A(NSUBT) = NCELL (1)
      B(NSUBT) = NCELL (2)
      C(NSUBT) = NCELL (3)
      D(NSUBT) = NCELL (4)
      NSUBT = NSUBT + 1
      NTOT = 0
      GO TO 2
 1000 CONTINUE
      IF (NLEVEL.EQ.12) NTST = 1
      NTST = 0
   23 CONTINUE
      NSUBT = NSUBT - 1
      DO 27 I = 1,NSUBT
C  FLAG ZERO ELEMENT TABLES
      TEST = A(I)*B(I)*C(I)*D(I)
      IF (TEST.EQ.ZERO) FLAG(NRUN) = STARS
   27 CONTINUE
      DO 24 I = 1,NSUBT
      IF (A(I).EQ.ZERO) A(I) = 0.1
      IF (B(I).EQ.ZERO) B(I) = 0.1
      IF (C(I).EQ.ZERO) C(I) = 0.1
      IF (D(I).EQ.ZERO) D(I) = 0.1
   24 CONTINUE
      DO 25 I = 1,NSUBT
C  FILL UP THE  OBS  TABLE FOR USE BY THE FIRST ORDER CHISQUARE  ROUTINE
       OBS (1,1,I) = A(I)
       OBS (2,1,I) = B(I)
       OBS (1,2,I) = C(I)
       OBS (2,2,I) = D(I)
C  SAVE OBS FOR LATER USE IN FILLING UP NEGATIVE CHISQUARE TABLES
      SAVE (1,I) = A(I)
      SAVE (2,I) = B(I)
      SAVE (3,I) = C(I)
      SAVE (4,I) = D(I)
   25 CONTINUE
C  DO NOT CALCULATE SECOND ORDER CHISQUARE IF ONE OR MORE ELEMENTS ARE ZERO
      IF (FLAG(NRUN).NE.STARS) GO TO 110
      CHISQ2(NRUN) = 0.0
      CHI2 = 0
      GO TO 111
  110 CONTINUE
```

```
C  CALL SUBROUTINE TO CALCULATE SECOND ORDER CHISQUARE
      CALL CHISEC (A,B,C,D,NSUBT,CHI2)
C  FILL UP SECOND ORDER CHISQUARE TABLE
      CHISQ2 (NRUN) = CHI2
  111 CONTINUE
      NT = NSUBT + 4
C  CALCULATIONS OF FIRST ORDER CHISQUARE
      DO 26 I = 1,NT
      TAU (I) = 0.0
   26 CONTINUE
      DO 30 I = 1,2
      DO 30 J = 1,NSUBT
      TAU (1) = TAU (1) +  OBS (1,I,J)
      TAU (2) = TAU (2) +  OBS (2,I,J)
      TAU (3) = TAU (3) +  OBS (I,1,J)
      TAU (4) = TAU (4) +  OBS (I,2,J)
   30 CONTINUE
      DO 35 I = 1,2
      DO 35 J = 1,2
      DO 35 K = 1,NSUBT
      TAU(K+4) = TAU(K+4) + OBS(I,J,K)
   35 CONTINUE
      CHISQ = 0.0
      TOT = TAU(1) + TAU(2)
      DO 45 I = 1,2
      DO 45 J = 1,2
      DO 45 K = 1,NSUBT
      THEORY       = (TAU(I)*TAU(J+2)*TAU(K+4))/(TOT*TOT)
      ACTUAL =  OBS (I,J,K)
      CHISQ = CHISQ + ((ACTUAL-THEORY)**2/THEORY)
   45 CONTINUE
C  END OF FIRST ORDER CHISQUARE CALCULATIONS
C  FILL UP SIGNIFICANT VALUE FREQUENCY TABLE FOR  FIRST& SECOND ORDER CHISQUARE
      IF (CHISQ.GE.CHITST(1)) NFREQ(1) = NFREQ(1) + 1
      IF (CHISQ.GE.CHITST(2)) NFREQ(2) = NFREQ(2) + 1
      IF (CHISQ.GE.CHITST(3)) NFREQ(3) = NFREQ(3) + 1
      IF (CHISQ.GE.CHITST(4)) NFREQ(4) = NFREQ(4) + 1
      IF (CHI2.GE.CHTST2(1)) NFRQ2(1) = NFRQ2(1) + 1
      IF (CHI2.GE.CHTST2(2)) NFRQ2(2) = NFRQ2(2) + 1
      IF (CHI2.GE.CHTST2(3)) NFRQ2(3) = NFRQ2(3) + 1
      IF (CHI2.GE.CHTST2(4)) NFRQ2(4) = NFRQ2(4) + 1
      CHISQ1 (NRUN) = CHISQ
C  CHECK FOR NEGATIVE CHISQUARES
      IF (CHISQ.LT.0.OR.CHI2.LT.0) GO TO 101
      GO TO 1100
  101 CONTINUE
      NEGATI = NEGATI + 1
C  FILL UP NEGATIVE CHISQUARE TABLE
      DO 104 I = 1,NSUBT
      NSAVE (NEGATI,1,I) = SAVE(1,I)
      NSAVE (NEGATI,2,I) = SAVE(2,I)
```

```
      NSAVE (NEGATI,3,I) = SAVE(3,I)
      NSAVE (NEGATI,4,I) = SAVE(4,I)
  104 CONTINUE
C  GO BACK & GENERATE ANOTHER TABLE
 1100 CONTINUE
      WRITE (3,891)
C  PRINT CONTROL DATA
      WRITE (3,894) (NI(I),I = 1,NLEVEL),NRUNS,IX
      WRITE (3,886)
      DO 105 I = 1,4
C  PRINT SIGNIFICANT VALUE FREQUENCY TABLES
  105 WRITE (3,885) CHITST(I),NFREQ(I),CHTST2(I),NFRQ2(I)
      IF (NEGATI.LE.0) GO TO 102
C  PRINT NEGATIVE CHISQUARES
      WRITE (3,888)
      DO 103 I = 1,NEGATI
      DO 103 K = 1,NSUBT
      WRITE (3,887) (NSAVE(I,J,K),J = 1,4)
  103 CONTINUE
  102 CONTINUE
      WRITE (3,893)
      NLINES = 0
      SIGX = 0.0
      SIGY = 0.0
      SIGXY = 0.0
      SIGXSQ = 0.0
      SIGYSQ = 0.0
      NRNS = 0
      DO 50 I = 1,NRUNS
      IF (NLINES.GE.45) GO TO 46
   47 CONTINUE
C  PRINT CHISQUARE TABLE
      WRITE (3,892) I,CHISQ1 (I),CHISQ2 (I),FLAG(I)
      NLINES = NLINES + 1
      IF (FLAG(I).NE.BL) GO TO 50
C  PUNCH CHISQUARES IF NECESSARY
      IF (NPUNCH.EQ.1) WRITE (2,897) CHISQ1(I),CHISQ2(I)
      NRNS = NRNS + 1
      SIGX = SIGX + CHISQ1(I)
      SIGY = SIGY + CHISQ2(I)
      SIGXY = SIGXY + CHISQ1(I)*CHISQ2(I)
      SIGXSQ = SIGXSQ + CHISQ1(I)*CHISQ1(I)
      SIGYSQ = SIGYSQ + CHISQ2(I)*CHISQ2(I)
   50 CONTINUE
      R = (NRNS *SIGXY - SIGX*SIGY)/(SQRT((NRNS *SIGXSQ - SIGX*SIGX) *
     1(NRNS *SIGYSQ - SIGY*SIGY)))
      WRITE (3,890) R
      SX = (SQRT(NRNS*SIGXSQ - SIGX*SIGX))/NRNS
      SY = (SQRT(NRNS*SIGYSQ - SIGY*SIGY))/NRNS
      SYX = SY*SQRT(1 - R*R)
      XBAR = SIGX/NRNS
```

```
      YBAR = SIGY/NRNS
      WRITE (3,896) SX,SY,SYX,SIGX,SIGY,SIGXY,SIGXSQ,SIGYSQ,XBAR,YBAR
      GO TO 1200
   46 WRITE (3,893)
      NLINES = 0
      GO TO 47
 1200 CONTINUE
      GO TO 3
 2000 CONTINUE
C  READ FORMATS
  999 FORMAT (I2,I4,I1,I6)
  998 FORMAT (12I3)
  997 FORMAT (12F5.2)
C PRINT FORMATS
  899 FORMAT (/,'  SUB TABLE NUMBER ',I4)
  898 FORMAT (7X,5I4)
  897 FORMAT (12F6.2)
  896 FORMAT (' SX =',F8.4,' SY =',F8.4,' SXY =',F8.4,' SIGX =',F8.2,' S
     1IGY =',F8.2,' SIGXY =',F10.2,' SIGXSQ =',F10.2,' SIGYSQ =',F10.2,
     2//,' XBAR =',F8.4,' YBAR =',F8.4)
  895 FORMAT (7X,5F5.0)
  894 FORMAT (/////,9X,' EXPECTED VALUE FOR LEVEL',15X,'= OF  SEED',/,
     12X,'1',3X,'2',3X,'3',3X,'4',3X,'5',3X,'6',3X,'7',3X,'8',3X,'9',2X,
     2'10',2X,'11',2X,'12',2X,'RUNS'2X,'VALUE',//,1X,12(I2,2X),I4,I6)
  893 FORMAT ('1',///,7X,' TABLE NUMBER',2X,'CHISQ-1',2X,'CHISQ-2',
     12X,'FLAG')
  892 FORMAT (7X,I8,7X,F6.2,2X,F8.2,3X,A2)
  891 FORMAT ('1')
  890 FORMAT (//,7X,'CORRELATION COEFFICIENT = ',F8.4)
  889 FORMAT ('1',///,7X,'***  NUMBER OF TABLES TO BE GENERATED IS MORE
     1 THAN 400. JOB CANCELLED  ***')
  888 FORMAT (/,7X,' TABLE OF NEGATIVE CHI-SQUARES')
  887 FORMAT (7X,5I20)
  886 FORMAT (//,7X,' DISTRIBUTION   OF   CHISQUARED VALUES',//,7X,'VAL
     1UE',2X,'FREQ',2X,'VALUE',2X,'FREQ',/)
  885 FORMAT (7X,2(F5.2,3X,I3,1X))
      STOP
      END
C  THIS IS THE RANDOM NUMBER GENERATOR USED IN GENERATING THE CONTINGENCY
C  TABLE. FOR MORE INFORMATION ON THIS METHOD REFER TO IBM MANUAL 020-8011,
C  RANDOM NUMBER GENERATION AND TESTING
      SUBROUTINE RANDU (IX,IY,NUM)
      IY = IY*IX
      IF (IY.LT.0) IY = IY + 2147483647 + 1
      RNUM = IY*0.4656613E-9
      NUM = RNUM*10.0
      RETURN
      END
C  THIS SUBROUTINE  IS A PROGRAM  OF 'NORTON'S' ROUTINE REFO
C  JOURNAL OF AMERICAN STATISTICAL ASSOCIATION, 1945,40,P251-258
      SUBROUTINE CHISEC (A,B,C,D,N,CHI2)
```

```
      DIMENSION A(1), B(1), C(1), D(1)
      DIMENSION P(12),S(12),X(12),ER(12),A1(12),A2(12),A3(12),A4(12),
     1 XX(24),CHI(12)
      INTEGER A1,A2,A3,A4
      K=0
      CHI2=0.0
      E = 0.003
      DO 10 I = 1,N
      X(I) = 0.0
      XX(I) = 0.0
      A1(I) = A(I)
      A2(I) = B(I)
      A3(I) = C(I)
      A4(I) = D(I)
 10   CONTINUE
 11   K = K + 1
      DO 15 I = 1,N
      A(I) = A(I)+X(I)
      B(I) = B(I)-X(I)
      C(I) = C(I)-X(I)
      D(I) = D(I)+X(I)
      P(I) = (A(I)*D(I))/(B(I)*C(I))
      Y=(B(I)-X(I))*(A(I)+X(I))*(D(I)+X(I))*(C(I)-X(I))
      R=(A(I)+X(I))*(D(I)+X(I))*(C(I)-X(I))
      RR=(B(I)-X(I))*(D(I)+X(I))*(C(I)-X(I))
      RS=(B(I)-X(I))*(A(I)+X(I))*(C(I)-X(I))
      RT=(B(I)-X(I))*(A(I)+X(I))*(D(I)+X(I))
      S(I)=Y/(R+RR+RS+RT)
 15   CONTINUE
      SUM =0.0
      SP = 0.0
      DO 20 I = 1,N
      SUM = SUM + S(I)
      SP = SP + S(I)*P(I)
 20   CONTINUE
      H=SUM/SP
      DO 25 I = 1,N
      X(I)= S(I)*(1.0-H*P(I))
      XX(I) = XX(I)+X(I)
 25   CONTINUE
      DO 30 I=1,N
      IF(ABS(X(I)).GT.E) GO TO 11
 30   CONTINUE
      IF(K.EQ.10) GO TO 65
      DO 45 I=1,N
      ER(I)=1.0/A(I)+1.0/B(I)+1.0/C(I)+1.0/D(I)
      CHI(I)= XX(I)**2*ER(I)
      CHI2=CHI2+CHI(I)
 45   CONTINUE
 65   CONTINUE
      RETURN
      END
```

APPENDIX II

Program For Second Order Interaction Chi-Square

```
C  THIS IS A PROGRAM OF 'NORTON'S' ROUTINE REFO
C  JOURNAL OF AMERICAN STATISTICAL ASSOCIATION, 1945,40,P251-258
      DIMENSION A(50),B(50),C(50),D(50),P(50),S(50),X(50),ER(50),A1(50),
     1A2(50),A3(50),A4(50),XX(53),CHI(50)
      INTEGER A1,A2,A3,A4
C  READ  ACCURACY DESIRED AND NUMBER OF 2X2 LEVELS
100   READ(1,1,END=1000)E,N
1     FORMAT(F5.3,5X,I3)
      K=0
      CHI2=0.0
      DO 5 I = 1,N
      READ(1,2) A(I), B(I), C(I), D(I)
C  READ CELL SIZES FOR EACH LEVEL
2     FORMAT(4F10.1)
5     CONTINUE
      DO 10 I = 1,N
      X(I) = 0.0
      XX(I) = 0.0
      A1(I) = A(I)
      A2(I) = B(I)
      A3(I) = C(I)
      A4(I) = D(I)
10    CONTINUE
11    K = K + 1
      DO 15 I = 1,N
      A(I) = A(I)+X(I)
      B(I) = B(I)-X(I)
      C(I) = C(I)-X(I)
      D(I) = D(I)+X(I)
      P(I) = (A(I)*D(I))/(B(I)*C(I))
      Y=(B(I)-X(I))*(A(I)+X(I))*(D(I)+X(I))*(C(I)-X(I))
      R=(A(I)+X(I))*(D(I)+X(I))*(C(I)-X(I))
      RR=(B(I)-X(I))*(D(I)+X(I))*(C(I)-X(I))
      RS=(B(I)-X(I))*(A(I)+X(I))*(C(I)-X(I))
      RT=(B(I)-X(I))*(A(I)+X(I))*(D(I)+X(I))
      S(I)=Y/(R+RR+RS+RT)
15    CONTINUE
      SUM =0.0
      SP = 0.0
      DO 20 I = 1,N
      SUM = SUM + S(I)
      SP = SP + S(I)*P(I)
20    CONTINUE
      H=SUM/SP
      DO 25 I = 1,N
      X(I)= S(I)*(1.0-H*P(I))
      XX(I) = XX(I)+X(I)
25    CONTINUE
      DO 30 I=1,N
      IF(ABS(X(I)).GT.E) GO TO 11
30    CONTINUE
```

```
      IF(K.EQ.10) GO TO 65
      DO 45 I=1,N
      ER(I)=1.0/A(I)+1.0/B(I)+1.0/C(I)+1.0/D(I)
      CHI(I)= XX(I)**2*ER(I)
      CHI2=CHI2+CHI(I)
 45   CONTINUE
      WRITE(3,50)
 50   FORMAT('1',////,7X,'LEVEL',7X,'A',4X,'B',4X,'C',4X,'D',22X,'XX(I)'
     1,10X,'ER(I)',10X,'CHI(I)',//)
      DO 56 I=1,N
      WRITE(3,55) I,A1(I),A2(I),A3(I),A4(I),XX(I),ER(I),CHI(I)
 55   FORMAT(9X,I2,5X,I4,1X,I4,1X,I4,1X,I4,20X,3(F15.7,5X))
 56   CONTINUE
      WRITE(3,60) K,E,CHI2
 60   FORMAT(///,5X,I2,1X,'APPROXIMATIONS,SUCH THAT THE CORRECTIONS DO N
     1OT EXCEED',1X,F5.3,1X,',THE CHI SQUARE HAS AVALUE OF',1X,F15.7)
 65   CONTINUE
      GO TO 100
1000  STOP
      END
```

APPENDIX III

Computer Output Sample With Associated Computations

EXPECTED VALUE FOR LEVEL												= OF	SEED
1	2	3	4	5	6	7	8	9	10	11	12	RUNS	VALUE
5	8	8	11	**	**								

TABLE OF NEGATIVE CHI-SQUARES

	4	3	3	10
3.58	8	7	12	5
	4	9	7	12
	10	15	9	10
	4	4	3	9
5.68	11	7	10	4
	3	10	9	10
	7	17	13	7

10	5	2	1
34	19	11	5
41	19	9	4

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
1	12.11	3.19	
2	10.62	5.79	
3	8.57	2.16	
4	12.85	2.54	
5	9.04	3.00	
6	5.68	3.53	
7	9.56	0.68	
8	10.12	5.39	
9	13.50	0.94	
10	7.26	0.67	
11	13.10	4.30	
12	11.28	7.84	
13	9.64	2.43	
14	7.01	5.31	
15	12.78	0.53	
16	7.70	3.72	
17	10.16	2.01	
18	12.46	8.66	
19	9.38	2.32	
20	12.32	2.12	
21	8.53	1.98	
22	13.79	3.34	
23	6.47	0.92	
24	7.38	3.02	
25	13.38	2.59	
26	8.71	3.87	
27	14.82	3.09	
28	15.89	2.04	
29	6.57	1.21	
30	5.65	2.52	
31	10.22	4.75	
32	7.42	5.23	
33	15.70	4.24	
34	19.07	7.69	
35	14.20	1.41	
36	6.12	0.62	
37	9.83	1.66	
38	7.86	1.41	
39	6.26	0.99	
40	18.46	2.26	
41	4.45	1.12	
42	7.03	2.16	
43	16.57	7.34	
44	12.32	3.97	
45	8.81	6.62	

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
46	3.91	0.66	
47	9.41	5.96	
48	3.96	3.21	
49	11.36	5.28	
50	9.04	3.09	
51	7.48	3.01	
52	13.32	1.31	
53	15.31	3.34	
54	14.43	3.26	
55	5.84	2.14	
56	6.59	0.66	
57	13.47	1.45	
58	10.81	4.91	
59	10.92	0.0	**
60	8.76	1.58	
61	10.45	7.18 ✓	
62	7.52	3.21	
63	12.34	6.67 ✓	
64	13.40	0.85	
65	2.44	0.26	
66	11.23	0.75	
67	9.29	5.56	
68	11.57	3.73	
69	11.01	6.44 ✓	
70	3.40	0.74	
71	14.90	2.43	
72	12.57	2.13	
73	10.83	5.13	
74	2.88	0.39	
75	5.18	2.11	
76	8.00	1.23	
77	11.78	1.08	
78	3.65	1.51	
79	8.63	3.64	
80	13.50	~~28.57~~ 3.58	
81	11.94	4.97	
82	6.22	2.34	
83	9.18	2.42	
84	12.97	1.24	
85	5.26	2.72	
86	7.91	1.56	
87	6.91	1.57	
88	5.94	1.33	
89	10.03	5.04	
90	10.39	1.96	

10 5 2 1

3 2 0 0

8 2 0 0

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
91	15.19	1.70	
92	14.14	3.27	
93	21.35	10.95	
94	11.71	4.26	
95	5.40	2.95	
96	4.13	1.70	
97	7.74	1.82	
98	8.43	3.32	
99	12.60	5.55	
100	13.13	3.34	
101	9.37	1.15	
102	17.07	1.21	
103	10.03	4.35	
104	3.80	1.17	
105	6.01	2.53	
106	5.13	1.76	
107	3.48	1.83	
108	9.33	5.69	
109	9.99	3.46	
110	4.76	1.62	
111	7.98	1.53	
112	9.30	4.95	
113	6.71	2.85	
114	7.99	3.89	
115	20.91	1.33	
116	12.84	3.60	
117	5.54	2.60	
118	19.45	15.71	
119	8.70	2.29	
120	9.56	4.66	
121	20.19	8.12	
122	9.54	0.98	
123	4.20	2.52	
124	10.82	3.49	
125	6.87	2.09	
126	11.52	0.82	
127	12.71	7.70	
128	30.41	12.30	
129	13.85	5.98	
130	13.70	3.30	
131	7.33	0.87	
132	11.29	2.29	
133	9.08	1.95	
134	10.39	1.41	
135	10.17	2.86	

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
136	14.54	1.80	
137	9.21	6.63	
138	6.27	0.96	
139	12.05	1.34	
140	16.49	9.88	
141	4.66	1.40	
142	5.77	1.30	
143	4.57	0.52	
144	3.11	0.50	
145	4.65	2.35	
146	8.42	0.49	
147	9.42	2.92	
148	7.10	3.73	
149	11.84	1.63	
150	11.26	1.27	
151	8.97	1.48	
152	8.79	0.95	
153	8.79	1.93	
154	13.99	2.02	
155	9.16	3.88	
156	7.57	2.19	
157	10.26	8.72	
158	12.88	6.34	
159	8.97	1.74	
160	12.93	2.03	
161	5.96	1.52	
162	7.04	4.82	
163	8.32	1.82	
164	23.84	5.23	
165	6.65	1.92	
166	6.27	2.62	
167	6.36	0.90	
168	5.45	0.95	
169	8.69	1.11	
170	9.62	1.29	
171	11.77	1.27	
172	7.38	2.63	
173	8.60	3.13	
174	10.26	0.39	
175	22.99	4.03	
176	8.06	2.27	
177	6.37	2.26	
178	9.87	2.05	
179	14.78	3.21	
180	10.77	2.97	

10 5 2 1
9 7 . 4 9
9 6 4 2

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
181	8.24	4.97	
182	8.05	2.11	
183	5.02	0.54	
184	5.90	4.09	
185	9.45	4.44	
186	10.82	3.59	
187	16.42	6.50	
188	9.41	3.67	
189	6.01	[illegible].23	
190	8.41	5.18	
191	18.45	13.04	
192	15.68	8.98	
193	14.85	3.81	
194	8.04	1.82	
195	6.42	3.16	
196	12.22	3.28	
197	20.70	7.16	
198	16.58	1.62	
199	17.69	7.96	
200	4.56	2.23	
201	6.89	1.63	
202	10.98	2.21	
203	7.00	1.24	
204	5.75	0.59	
205	6.51	3.28	
206	3.20	0.42	
207	16.33	1.97	
208	5.90	0.03	
209	8.22	1.17	
210	26.19	0.0	**10,5
211	6.68	3.22	
212	17.84	5.95	
213	13.18	2.05	
214	6.84	3.57	
215	11.09	6.61	
216	10.42	1.40	
217	15.66	1.03	
218	8.91	4.67	
219	6.27	1.07	
220	12.12	1.52	
221	8.71	2.13	
222	3.52	0.31	
223	9.52	6.06	
224	7.93	1.10	
225	12.91	1.97	

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
226	3.23	0.53	
227	11.53	7.13	
228	5.18	2.23	
229	5.20	2.30	
230	7.67	5.23	
231	3.92	1.87	
232	7.49	0.67	
233	22.49	17.98	
234	6.01	0.63	
235	6.33	2.60	
236	24.89	10.60	
237	7.66	2.57	
238	16.17	4.74	
239	3.67	0.35	
240	5.61	0.31	
241	3.66	0.72	
242	16.66	2.93	
243	15.30	8.14	
244	6.91	1.10	
245	13.17	1.25	
246	11.14	9.43	
247	5.81	2.67	
248	8.59	1.23	
249	15.05	0.0	**
250	15.26	6.25	
251	4.28	0.09	
252	6.55	1.26	
253	12.95	4.61	
254	12.30	3.35	
255	4.52	1.66	
256	8.44	1.71	
257	7.97	4.83	
258	7.59	4.10	
259	8.71	3.23	
260	6.90	2.43	
261	6.64	3.89	
262	15.79	6.84	
263	18.33	9.96	
264	6.02	1.12	
265	11.33	0.91	
266	5.78	1.92	
267	5.06	1.35	
268	8.01	1.32	
269	16.67	4.28	
270	10.10	2.69	

10 5 2 1
14 6 3 2
15 9 4 2

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
271	15.76	1.76	
272	2.53	0.34	
273	4.93	0.55	
274	16.60	1.98	
275	8.73	5.11	
276	7.80	0.89	
277	15.06	4.27	
278	6.95	1.13	
279	6.66	0.86	
280	6.74	2.31	
281	12.77	0.69	
282	15.05	2.56	
283	12.76	5.01	
284	3.78	0.73	
285	10.51	2.59	
286	10.24	3.76	
287	12.40	3.16	
288	7.89	1.43	
289	15.04	8.32	
290	6.67	1.67	
291	2.63	0.64	
292	9.96	1.00	
293	8.94	3.62	
294	15.36	2.64	
295	9.59	2.02	
296	8.00	0.32	
297	5.96	1.07	
298	10.42	0.25	
299	6.18	1.62	
300	8.44	4.21	
301	5.86	5.05	
302	9.41	5.11	
303	4.83	1.74	
304	5.79	3.61	
305	9.59	1.43	
306	4.96	2.41	
307	8.98	2.08	
308	7.62	2.26	
309	9.77	4.91	
310	15.19	2.68	
311	7.40	3.64	
312	6.41	3.61	
313	23.33	7.16	
314	13.04	3.57	
315	8.64	1.64	

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
316	3.73	0.19	
317	8.45	3.61	
318	5.03	3.89	
319	22.51	4.97	
320	10.83	4.15	
321	12.84	4.32	
322	12.19	1.40	
323	10.37	1.41	
324	6.32	2.46	
325	12.97	3.56	
326	12.52	2.38	
327	7.58	2.53	
328	11.66	6.27	
329	13.56	2.77	
330	11.29	1.03	
331	13.64	7.56	
332	12.42	5.49	
333	5.31	1.37	
334	14.86	2.77	
335	8.58	4.02	
336	1.66	0.41	
337	9.31	1.89	
338	11.02	0.72	
339	5.06	2.62	
340	16.11	11.30	
341	6.25	0.08	
342	8.74	1.43	
343	10.97	4.80	
344	12.57	2.97	
345	4.54	2.35	
346	5.46	0.68	
347	5.91	0.93	
348	13.77	0.0	**
349	4.85	0.96	
350	12.90	2.16	
351	13.09	7.14	
352	4.11	2.21	
353	12.21	7.20	
354	7.17	1.07	
355	8.81	6.14	
356	7.97	1.15	
357	5.81	1.05	
358	9.82	1.03	
359	6.03	0.81	
360	10.18	2.24	

4 2 2 1
7 2 1 0

TABLE NUMBER	CHISQ-1	CHISQ-2	FLAG
361	11.35	4.20	
362	13.80	1.33	
363	11.74	1.35	
364	8.17	4.08	
365	12.29	4.29	
366	8.12	2.07	
367	5.02	3.45	
368	4.49	0.53	
369	11.95	1.96	
370	10.20	6.76	
371	3.70	0.63	
372	16.60	2.50	
373	9.26	1.16	
374	3.88	0.98	
375	10.72	1.11	
376	21.37	3.07	
377	23.19	6.18	
378	6.56	2.56	
379	6.72	0.30	
380	9.47	0.77	
381	13.25	3.68	
382	15.14	1.43	
383	9.05	4.88	
384	18.26	~~1.36~~ 5.68	
385	5.51	3.31	
386	3.97	0.37	
387	10.65	3.38	
388	6.81	0.87	
389	6.32	0.12	
390	5.23	0.86	
391	8.11	2.72	
392	14.49	1.35	
393	10.78	0.0	
394	7.82	0.60	
395	3.36	0.59	
396	3.80	0.94	
397	11.22	1.98	
398	11.82	4.02	
399	13.85	6.99	
400	3.76	0.71	

10 5 2 1
4 9 2 0
2 0 0 0

CORRELATION COEFFICIENT = ~~0.2960~~ 0.5627

SX = 4.4061 SY = ~~3.6705~~ 2.7593 SXY = 3.5060 SIGX = 3890.68 SIGY = ~~1105.24~~ 1184.73 SIGXY = ~~12754.79~~ 14047.77 SIGXSQ = 45913.58 SIGYSQ = ~~8419.85~~ 5938.03

XBAR = 9.8249 YBAR = ~~2.7910~~ 2.9909

The program for obtaining the results of the Monte Carlo study was carefully checked by hand results as explained elsewhere. However, due to the necessity of obtaining the results quickly while the computer was available, no time was spent on making the computer output pleasing to the eye. One might say "utility" was the motto.

The previous ten pages show how the majority of the results came from the computer. Toward the end of the study steps were incorporated in the program to sum up the values of the two columns greater than the values of the tabulated chi-square for the 10%, 5%, 2% and 1% levels of significance.

The contingency table under consideration was a 2 x 2 x 4 table with subtable sample sizes of 20, 32, 32 and 44. Tables 80 and 384 have negative values for second order chi-square. These tables were printed out. They were subjected to the Con Midhe shift and recomputed using the computer program of Appendix II. The new values are written in along side of the negative value. Two stars in the column marked "flag" indicate a contingency table with a zero cell. In order to avoid confusion the chi-square for second order interaction is then just printed as 0.0. For this computer run there are four such tables 59, 210, 249 and 348. In order to determine their predicted value by using regression analysis and the first order interaction, some corrections had to be made to the calculations on the last page of the printout. First, a translation is in order:

Printout	Text
SX	S_x
SY	S_y
SXY	S_{xy}
SIGX	X
SIGY	Y
SIGXY	XY
SIGXSQ	X^2
SIGYSQ	Y^2
XBAR	$\overline{X}$
YBAR	$\overline{Y}$

Since in this computer run there were negative chi-squares the following items were in error: correlation coefficients, SY, SIGY, SIGXY, SIGYSQ, and YBAR. These were corrected by means of a Victor electronic calculator. What was done then was to calculate the required X to obtain a Y equal to 6.25, 7.82, 9.84, or 11.35, the 10%, 5%, 2%, and 1% levels of significance for the second order interaction of a 2 x 2 x 4 contingency table. In this way only four calculations were necessary for any number of tables with zero valued cells. For table 210 the value of the first order chi-square was such as to predict a value for the second order chi-square greater than the minimum value for 5% level of significance but less than the value for the 2% level of significance, so 10, 5 was written next to the stars meaning a count for both the 10% level and 5% level. Next to tables 87, 177, 267, 357 and 396 there are the written numbers 10, 5, 2, and 1 with two rows of numbers beneath them. The first row is for the number of tables that exceeded or equaled the 10%, 5% and 2%, and 1% level for the first order chi-square and the second row is for the corresponding values for the second order chi-square. For this run there were the following values:

level	10%	5%	2%	1%
first order chi-square	34	19	11	5
second order chi-square	41	19	9	4

These are then converted to the following percentages:

level	10%	5%	2%	1%
first order chi-square	8.5%	4.75%	2.75%	1.25%
second order chi-square	10.25%	4.75%	2.25%	1.00%

These values were then recorded in the table in Chapter Four along with the value of 1% in the column labeled "% Tables w/zero cells".

BIBLIOGRAPHY

Atthowe, J. M. Behavior Innovation and Persistance. American Psychologist, 1973, 28, pp. 34-42.

Attneave, Fred. Applications of Information Theory to Psychology. Henry Holt and Co., New York, 1959.

Bartlett, M. S. "Contingency Table Interactions". Journal of Royal Statistical Society Supplement, 1935, 2, pp. 248-252.

Barton, Richard D. A Primer on Simulation and Gaming. Prentice-Hall Inc., 1970, Englewood Cliffs, New Jersey.

Berkson, J. "Application of minimum logit chi-square estimate to a problem of Grizzle with a notation on the problem of 'no interaction'". Biometrics, 1968, 24, pp. 75-98.

Bhapkar, V. P. "Some Tests for Categorical Data". Annals of Mathematical Statistics, 1961, 32, pp. 72-83.

Bhapkar, V. P. and Koch, Gary G. "Hypotheses of 'No Interaction' in Multidimensional Contingency Tables". Technometrics, 1968, 10, pp. 107-123.

________. "On the Hypotheses of 'No interaction' in Contingency Tables". Biometrics, 1968, 24, pp. 567-594.

Birch, M. W. "Maximum Likelihood in Three-Way Contingency Tables". Journal of Royal Statistical Society, 1963, 25B, pp. 220-233.

Bishop, Yvonne M. M. "Effects of Collapsing Multidimensional Contingency Tables". Biometrics, 1971, 25, pp. 545-562.

Boyd, William C. "A Nomogram For Chi-Square". American Statistical Association Journal, 1965, 61, pp. 344-346.

Camp, B. H. "Further Interpretations of the Chi-Square Test". Journal of American Statistical Association, 1938, 33, pp. 537-542.

Childs, G. B. "An Analysis of the Success in Initial University Mathematics Courses of Students with Correspondence Study and Non-correspondence Study Backgrounds in High School". Journal of Educational Research, 1956, 49, pp. 606-610.

________. "Correspondence Study: "Concepts and Comments" Address at the annual meeting of the National University Extension Association, April 13, 1973.

Cochran, W. G. "The Chi-square Test of Goodness of Fit". Annals of Mathematical Statistics, 1952, 23, pp. 315-345.

________. "Some Methods for Strengthening the Common Chi-Square Tests". Biometrics, 1954, 10, pp. 417-451.

Cohan, A. S. Career Patterns in the Irish Political Elite. British Journal of Political Science, 1973, 3, pp. 213-228.

Conover, W. J. Practical Nonparametric Statistics. Wiley, New York, 1971.

Corsten, L. C. A. "Partition of Experimental Vectors Connected With Multinomial Distributions". Biometrics, 1957, 13, pp. 451-484.

Craddock, J. M., and Flood, C. R. "The Distribution of the Chi-Square Statistic in Small Contingency Tables". Applied Statistics Journal of Royal Statistical Society, 1970, 19C, pp. 173-181.

Darroch, J. M. "Interaction in Multi-Factor Contingency Tables". Journal of Royal Statistical Society, 1962, 24B, pp. 251-263.

Deming, W. E. and Stephan, F. F. "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known". Annals of Mathematical Statistics, 1940, 11, pp. 427-444.

Diamond, E. L. and Lilienfeld, A. M. "Effects of errors in classification and diagnosis in various types of epidemiological studies". American Journal of Public Health, 1962, 52, pp. 1137-1144.

Diamond, E. L., Mitra, S. K., and Roy, S. N. "Asymptotic power and asymptotic independence in the statistical analysis of categorical data". Bulletin of the Institute of International Statistics, 1960, 37, pp. 309-329.

Duncan, A. J. "Chi-Square Test of Independence and the Comparison of Percentages". Industrial Quality Control, 1955, 11, pp. 9-14.

Eagleson, G. K. "Polynomial Expansions of Bivariate Distributions". Annals of Mathematical Statistics, 1964, 35, pp. 1208-1215.

Feller, W. K. "Statistical Aspects of E.S.P.". Journal of Parapsychology, 1940, 4, pp. 271-298.

Fisher, R. A. Statistical Methods For Research Workers. Oliver and Boyd, Edinburgh, 1936.

Fisher, R. A. and Yates, F. Statistical Tables for Biological, Agricultural, and Medical Research. Oliver and Boyd, Edinburgh, 1943.

Fisher, R. A. Contributions to Mathematical Statistics. J. Wiley and Sons, New York, 1950.

Freeman, G. H. and Halton, J. H. "Note on an Exact Treatment of Contingency Goodness of Fit and Other Problems of Significance". Biometrics, 1951, 38, pp. 141-149.

Fricke, B. G. A Configural-Content-Intensity Item for Personality Measurement. Educational and Psychological Measurement, 1956, 16, pp. 54-62.

Garner, W. R. and McGill, W. J. "The Relation Between Information and Variance Analyses". Psychometrics, 1956, 21, pp. 219-228.

Gart, J. J. "Alternative Analyses of Contingency Tables". Journal of Royal Statistical Society, 1966, 28B, pp. 164-179.

Goldberg, H. and Levine, H. "Approximate formulas for the percentage points and normalization of t and chi-square". Annals of Mathematical Statistics, 1946, 17, pp. 216-225.

Good, I. J. "On the Estimation of Small Frequencies in Contingency Tables". Journal of Royal Statistical Society 1956, 18B, pp. 113-124.

________. "The Interaction Algorithm and Practical Fourier Analysis". Journal of Royal Statistical Society, 1958, 20B, pp. 361-372.

________. "The Interaction Algorithm and Practical Fourier Analysis: an addendum". Journal of Royal Statistical Society, 1960, 22B, pp. 372-375.

________. "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables". Annals of Mathematical Statistics, 1963, 34, p. 922.

Goodman, L. A. and Kruskal, W. H. "Measures of Association for Cross-Classifications". Journal of American Statistical Association, 1954, 49, pp. 732-764.

________. "Measures of Association for Cross-Classifications; II: Further Discussion and References". Journal of American Statistical Association, 1951, 54, pp. 123-163.

Goodman, L. A. "On Plackett's Test for Contingency Table Interactions". Journal of Royal Statistical Society, 1963, 25B, pp. 179-188.

________. "Simple Methods for Analyzing Three-Factor Interaction in Contingency Tables". Journal of American Statistical Association, 1964, 59, pp. 319-351.

Gray, J. and Tocher, J. F. "The Physical Characteristics of Adults and School Children in East Aberdeenshire". Journal of Anthropological Institute, 1900, 30, pp. 104-124.

Greeley, A. M. The Religious Factor and Academic Careers. American Journal of Sociology, 1973, 78, pp. 1247-1256.

Greenberg, R. A. and White, C. "The detection of a correlation between the sexes of adjacent sibs in human families". Journal of American Statistical Association, 1965, 60, pp. 1035-1045.

Gumbel, E. J. "On the reliability of the classical chi-square test". Annals of Mathematical Statistics, 1943, 14, pp. 253-263.

Haldane, J. B. S. "Substitutes for Chi-Square". Biometrics, 1955, 42, pp. 265-266.

Hayashi, Takakura, Makita, Saito. "One method of quantification of attitude II Analysis of political attitude". (Japanese, English summary.) Procedings of the Institute of Statistical Mathematics, 1958, 1960, 6, 8, pp. 1-39 and 45-46.

Haynam, G. E. and Leone, F. C. "Analysis of Categorical Data". Biometrics, 1965, 52, pp. 654-660.

Hoel, P. G. "On the Chi-Square Test for Small Samples". Annals of Mathematical Statistics, 1938, 9, pp. 158-165.

Hoyt, C. J., Krishnaiah, P. R. and Torrance, E. P. "Analysis of Complex Contingency Data". Journal of Experimental Education, 1959, 27, pp. 187-194.

Kastenbaum, M. A. and Lamphiear, D. E. "Calculation of Chi-Square to Test the No Three Factor Interaction Hypothesis". Biometrics, 1959, 15, pp. 107-115.

Kastenbaum, M. A. "A note on the additive partitioning of chi-square in contingency tables". Biometrics, 1960, 16, pp. 416-422.

Kavanagh, J. A. A Monte Carlo Study of the Polynomial Discriminant Method for Pattern Recognition. Unpublished Ph.D. Thesis, University of Minnesota, 1972.

Kendall, M. G. The Advanced Theory of Statistics. Volume I, Griffin and Co., London, 1947.

Kendall, M. G. and Stuart, A. The Advanced Theory of Statistics, 2, Inference and Relationship. Griffin and Co., London, 1961.

Kimball, A. W. "Short-cut formulas for the exact position of chi-square in contingency tables". Biometrics, 1954, 10, pp. 452-458.

Koch, Gary G. and Reinfurt, D. W. "The Analysis of Categorical Data from Mixed Models". Biometrics, 1971, 27, pp. 157-173.

Ku, H. H. "A Note on Contingency Tables Involving Zero Frequencies and the 2I Test". Technometrics, 1963, 5, pp. 398-400.

Ku, H. H. and Kullback, S. "Interaction in Multidimensional Contingency Tables: An Information Theoratic Approach". Journal of Research of National Bureau of Standards - Mathematical Sciences, 1968, 72B, No. 3, pp. 159-199.

Kullback, S. Information Theory and Statistics, Dover Publications, New York, 1968.

Lancaster, H. O. "Complex Contingency Tables Treated by the Partition of Chi-Square". Journal of the Royal Statistical Society, 1951, 13B, pp. 242-249.

________. The Chi-Square Distribution. Wiley, New York, 1969.

Larson, E. L. "Comparative Quality of Work Done by Students in Residence and in Correspondence Work". Journal of Educational Research, 1932, 25, pp. 105-109.

Lazarsfeld, P. F. "A Conceptual Introduction to Latent Structure Analysis". In P. F. Lazarsfeld (Ed.), Mathematical Thinking in the Social Science, 1954, Free Press, Glencoe, pp. 349-387.

Lewis D. and Burke, C. J. "The Use and Misuse of the Chi-Square Test". Psychological Bulletin, 1949, 46, pp. 433-489.

Lewis B. N. "On the Analysis of Interaction in Multidimensional Contingency Tables". Journal of the Royal Statistical Society, 1962, 125, pp. 85-95.

Lewis, D. and Burke, C. J. "Further Discussion of the Use and Misuse of the Chi-Square Test". Psychological Bulletin, 1950, 47, pp. 347-355.

Lewontin, R. C. and Felsenstein, J. "The Robustness of Homogeneity Tests in 2 x N Tables". Biometrics, 21, 1965, pp. 19-27.

Leyton, Morley K. "Rapid Calculation of Exact Probabilities for 2 x 3 Contingency Tables". Biometrics, 1968, 24, pp. 714-717.

Lindley, D. V. "The Bayesian Analysis of Contingency Tables". Annals of Mathematical Statistics, 1964, 35, pp.1622-1643.

Maxwell, A. E. Analysing Qualitative Data. John Wiley and Sons, New York, 1961.

Mayo, S. T. "Recent Techniques for Analyzing Association in Contingency Tables as Applied to an Analytical Follow-up Survey of Education Graduates". Journal of Experimental Education, 1957, 25, pp. 213-232.

Mayo, S. T. "Towards Strengthening the Contingency Table as a Statistical Method". Psychological Bulletin, 1959, 56, pp. 461-470.

________. "Interactions Among Categorical Variables". Educational and Psychological Measurement, 1961, 21, pp. 839-857.

Mitra, S. K. Contributions to the Statistical Analysis of Categorical Data. Institute of Statistics, University of North Carolina, Mimeograph Series No. 142, 1955.

Mood, A. M. Introduction to the Theory of Statistics. McGraw-Hill, Inc., New York, 1950, p. 280.

Morse, S. J. "Help, Likeability, and Social Influence". Journal of Applied Social Psychology, 1972, 2, pp. 24-34.

Mosteller, F. "Association and Estimation in Contingency Tables". Journal of American Statistical Association, 1968, 63, pp. 1-28.

Nachman, M. and Opochinsky, S. "The Effects of Different Teaching Methods: A Methodological Study". Journal of Educational Psychology, 1958, 49, pp. 245-249.

Naylor, Thomas H., et.al. Computer Simulation Techniques. John Wiley and Sons, New York, 1966.

Norton, H. W. "Calculation of Chi-Square For Complex Contingency Tables". Journal of American Statistical Association, 1945, 40, pp. 251-258.

Parsons, T. S. "A Comparison of Instruction by Kinescope, Correspondence Study and Customary Classroom Procedures", Journal of Educational Psychology, 1957, 48, pp. 27-40.

Patnaik, P. B. "The Non-Central Chi-Square and F-Distributions and Their Applications". Biometrics, 1949, 36, pp.202-232.

Peters, C. C. "The Misuse of Chi-Square - A Reply to Lewis and Burke". Psychological Bulletin, 1950, 47, pp. 331-337.

Pierce, Albert. Fundamentals of Nonparametric Statistics. McGraw Hill, New York, 1970.

Plackett, R. L. "A Note on Interactions in Contingency Tables". Journal of Royal Statistical Society, Series B, 1962, 24, pp. 162-166.

Pritsker, A. A. B. and Kiviat, P. J. Simulation With Gasp II. Prentice Hall, Englewood Cliffs, New Jersey, 1969.

Roy, S. N. and Bargmann, R. E. "Tests of Multiple Independence and the Associated Confidence Bounds". Annals of Mathematical Statistics, 1958, 29, pp. 491-503.

Roy, S. N. and Kastenbaum, M. A. "On the Hypothesis of No Interaction in a Multi-Way Contingency Table". Annals of Mathematical Statistics, 1956, 27, pp. 749-757.

________. A Generalization of Analysis of Variance and Multivariate Analysis to Data Based on Frequencies in Qualitative Categories or Class Intervals. Institute of Statistics, University of North Carolina, Mimeograph Series No. 131, 1955.

Roy, S. N. and Mitra, S. K. An Introduction to Some Nonparametric Generalizations of Analysis of Variance and Multivariate Analysis. Institute of Statistics, University of North Carolina, Mimeograph Series No. 139, 1955.

Roy, S. N. and Mitra, S. K. "An introduction to some non-parametric generalizations of analysis of variance and multivariate analysis". Biometrics, 1956, 43, pp. 361-376.

Shannon, C. E. and Weaver, W. The Mathematical Theory of Communication. University of Illinois Press, Urbana, 1963.

Simpson, E. H. "The Interpretation of Interaction in Contingency Tables". Journal of the Royal Statistical Society, Series B, 1951, 13, pp. 238-241.

Snedecor, G. W. "Chi-Squares of Bartlett, Mood and Lancaster, in a 2 x 2 x 2 Contingency Table". Biometrics, 1958, 14, pp. 560-562.

Somers, R. H. "Simple Meaaures of Association for the Triple Dichotomy". Journal of the Royal Statistical Society, 1964, A127, pp. 409-415.

Stanley, J. C. "Analysis of Unreplicated Three-Way Classifications with Applications to Rater Bias and Trait Independence". Psychometrics, 1961, 26, pp. 205-219.

Stuart, A. "The Estimation and Comparison of Strengths of Association in Contingency Tables". Biometrics, 1953, 40, pp. 105-110.

Sugiura, Nariaki & Otake, Masanori. "Numerical comparison of improved methods of testing in contingency tables with small frequencies". Annals of the Institute of Statistical Mathematics, 1968, 20, pp. 505-517.

Sutcliffe, J. P. "A general method of analysis of frequency data for multiple classification designs". Psychological Bulletin, 1957, 54, pp. 134-137.

Tate, M. W. Statistics in Education and Psychology. Macmillan, New York, 1965.

Walker, H. M. "Degrees of Freedom." Journal of Educational Psychology, 1940, 31, p. 253.

Weida, F. M. "On Measures of Contingency". Annals of Mathematical Statistics, 1934, 5, pp. 308-319.

Wilks, S. S. "The Likelihood Test of Independence in Contingency Tables". Annals of Mathematical Statistics, 1935, 6, pp. 190-196.

Yates, F. "Contingency Tables Involving Small Numbers of the Chi-Square Test". Journal of the Royal Statistical Society, Supplement 1, 1934, pp. 217-235.

Yule, G. U. "On the application of the chi-square method to association and contingency tables, with experimental illustrations". Journal of the Royal Statistical Society, 1922, 85, pp. 95-104.

Yule, G. U. and Kendall, M. G. An Introduction to the Theory of Statistics. Griffin and Co., London, 1947.

"Sirim dom hilluag mo saethir
A lenmain alt cen diehill
Cin neimnitnecht nacrad
Ocus atrab nid richith."

Book of Dimna, A.D. 620

"Would I might have as wages,
For work these pages given,
Freedom from critics scorning
And morning peace in heaven."

APPROVAL SHEET

The dissertation submitted by Raymond J. McNamee, has been read and approved by members of the Foundations Department.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

11/29/73
Date

Samuel T. Mayo
Signature of Advisor